University of Tennessee, Knoxville
Trace: Tennessee Research and Creative Exchange

Doctoral Dissertations | Graduate School

12-2013

# Virtual Reality Aided Mobile C-arm Positioning for Image-Guided Surgery

Zhenzhou Shao
zshao@utk.edu

This Dissertation is brought to you for free and open access by the Graduate School at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Zhenzhou Shao entitled "Virtual Reality Aided Mobile C-arm Positioning for Image-Guided Surgery." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biomedical Engineering.

Jindong Tan, Major Professor

We have read this dissertation and recommend its acceptance:

William R. Hamel, Xiaopeng Zhao, Hairong Qi

Accepted for the Council:
Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

# Virtual Reality Aided Mobile C-arm Positioning for Image-Guided Surgery

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Zhenzhou Shao

December 2013

© by Zhenzhou Shao, 2013
All Rights Reserved.

*DEDICATION*

*This dissertation is dedicated to my parents, Fengshi Li and Mingyu Shao, and my wife Ying Qu.*

# Acknowledgements

I would like to express my sincere and deep gratitude to my advisor, Dr. Jindong Tan, for his guidance and inspiration in the research and life. It was my great honor to join the research group under the supervision of Dr. Tan, without whose enlightening I would not have found a potential research career in biomedical field. I could never be able to convey my gratitude to Dr. Tan fully for his encouragement and enthusiasm for encountered problems in my research and life.

My deep appreciation also goes to Dr. William R. Hamel, Dr. Xiaopeng Zhao and Dr. Hairong Qi, who gave me valuable suggestions on my dissertation and served as my committee members of dissertation defense.

In my research, it was my honor to work with brilliant people who are generous with their time and help. Many thanks go to Professor Jianda Han and Professor Wei Liang for their generous support and help on experiments in Shenyang Institute of Automation (SIA), Chinese Academy of Sciences, China. Also thank their students Mr. Yang Li and Mr. Guoli Song, who worked closely with me and contributed much valuable solutions and experiments in the research work. I appreciate the generous help, encouragement and friendship from Dr. Sheng Hu, Dr. Lufeng Shi, Dr. Fanyu Kong, Dr. Shuo Huang, Dr. Xinying Zheng, Dr. Ya Tian and Mr. Jian Lu who were my labmates in the Robotics and Wireless Sensor Networks lab at Michigan Technological University. I also want to thank all my labmates both at Michigan Technological University and the University of Tennessee, Knoxville. We all experienced the long-distance moving from Houghton, MI to Knoxville, TN. Thanks

for their support and suggestions on my research and dissertation writing from 2012 to 2013: Dr. Hongsheng He, Dr. Xi Chen, Mr. Xiaolong Liu and Mr. Hao Zhang. I also would like to express my gratitude to Professor Rong Zhang, Professor Jie Zhang, Professor Binrui Wang and Professor Xiaoxian He for exchanges of knowledge and sharing of the rich experience.

Last but not least, I would like to specially thank all my family members for their consistent support, encouragement and love, who encourage me, support me and love through my life, without which I would not be who I am today.

# Abstract

Image-guided surgery (IGS) is the minimally invasive procedure based on the pre-operative volume in conjunction with intra-operative X-ray images which are commonly captured by mobile C-arms for the confirmation of surgical outcomes. Although currently some commercial navigation systems are employed, one critical issue of such systems is the neglect regarding the radiation exposure to the patient and surgeons. In practice, when one surgical stage is finished, several X-ray images have to be acquired repeatedly by the mobile C-arm to obtain the desired image. Excessive radiation exposure may increase the risk of some complications. Therefore, it is necessary to develop a positioning system for mobile C-arms, and achieve one-time imaging to avoid the additional radiation exposure.

In this dissertation, a mobile C-arm positioning system is proposed with the aid of virtual reality (VR). The surface model of patient is reconstructed by a camera mounted on the mobile C-arm. A novel registration method is proposed to align this model and pre-operative volume based on a tracker, so that surgeons can visualize the hidden anatomy directly from the outside view and determine a reference pose of C-arm. Considering the congested operating room, the C-arm is modeled as manipulator with a movable base to maneuver the image intensifier to the desired pose. In the registration procedure above, intensity-based 2D/3D registration is used to transform the pre-operative volume into the coordinate system of tracker. Although it provides a high accuracy, the small capture range hinders its clinical use due to the initial guess. To address such problem, a robust and fast initialization method is proposed

based on the automatic tracking based initialization and multi-resolution estimation in frequency domain. This hardware-software integrated approach provides almost optimal transformation parameters for intensity-based registration. To determine the pose of mobile C-arm, high-quality visualization is necessary to locate the pathology in the hidden anatomy. A novel dimensionality reduction method based on sparse representation is proposed for the design of multi-dimensional transfer function in direct volume rendering. It not only achieves the similar performance to the conventional methods, but also owns the capability to deal with the large data sets.

# Table of Contents


**1 Introduction** **1**

1.1 Mobile C-arm Positioning in Image-Guided Surgery . . . . . 1

1.2 Research Goals . . . . . 4

1.3 Research Challenges . . . . . 6

1.3.1 Mobile C-arm positioning . . . . . 6

1.3.2 Initialization of 2D/3D registration . . . . . 7

1.3.3 Visualization . . . . . 8

1.4 Contributions . . . . . 10

1.5 Dissertation Outline . . . . . 12

**2 Visual Reality Aided Positioning of Mobile C-arms** **13**

2.1 Introduction . . . . . 15

2.2 Materials . . . . . 16

2.2.1 Camera Augmented Mobile C-arm . . . . . 17

2.2.2 Vicon motion tracking system . . . . . 17

2.2.3 Spinal phantom . . . . . 18

2.3 VR Aided Determination of Reference Pose . . . . . 19

2.3.1 Reconstruction of Appearance Model . . . . . 19

2.3.2 Calibration of Appearance Model and Pre-operative CT . . . 20

2.3.3 Reference Pose Optimization of Mobile C-arm . . . . . 21

2.4 Kinematic Modeling and Control of the Mobile C-Arm . . . . . 25

2.4.1 Forward Kinematics of Mobile C-arm . . . . . 26
2.4.2 Inverse Kinematics with Base Movement Control . . . . . 29
2.5 Experimental Results and Discussions . . . . . 31
2.5.1 Accuracy of Appearance Model . . . . . 32
2.5.2 Simulation of reference pose determination . . . . . 32
2.5.3 Precision of Mobile C-arm positioning . . . . . 33
2.6 Related Work . . . . . 35
2.7 Summary . . . . . 37

**3 Robust and Fast Initialization for Intensity-based 2D/3D Registration 39**

3.1 Introduction . . . . . 41
3.2 Methods . . . . . 42
3.2.1 Level I: Tracking Based Initialization . . . . . 43
3.2.2 Level II: Multi-resolution Rotation Estimation and Translation Estimation in Fourier Domain . . . . . 45
3.3 Experiments and Discussions . . . . . 52
3.3.1 Performance Measurements . . . . . 53
3.3.2 Test of Multi-resolution Rotation Estimation . . . . . 54
3.3.3 Test of Complete Initialization . . . . . 56
3.3.4 Test of Intensity-based Registration Using Complete Initialization 58
3.4 Related Work . . . . . 59
3.4.1 C-arm Pose Tracking . . . . . 59
3.4.2 Pyramid Representations . . . . . 61
3.5 Discussion and Summary . . . . . 62

**4 Geometry Constrained Sparse Embedding for Multi-dimensional Transfer Function Design in Direct Volume Rendering 63**

4.1 Introduction . . . . . 65
4.2 Overview of Geometry Constrained Sparse Embedding . . . . . 67

4.3 Neighbor Selection Using Geometry Distance . . . . . 69
4.3.1 Geometry Distance . . . . . 69
4.3.2 Implementation of Neighbor Selection . . . . . 72
4.4 Dictionary Learning . . . . . 74
4.4.1 Objective Function . . . . . 74
4.4.2 Sparse Coding . . . . . 76
4.4.3 Dictionary Update . . . . . 77
4.5 Experimental Results and Discussions . . . . . 78
4.5.1 Comparison with Conventional Methods . . . . . 79
4.5.2 Comparison with Sparse Embedding . . . . . 81
4.5.3 Evaluation of Transfer Function in DVR . . . . . 83
4.6 Related Work . . . . . 85
4.6.1 Multi-dimensional Transfer Function . . . . . 85
4.6.2 Dimensionality Reduction and Sparse Representation . . . . . 87
4.7 Summary . . . . . 88

**5 Conclusions and Future Work** **89**
5.1 Conclusions . . . . . 89
5.2 Future Work . . . . . 91

**Bibliography** **94**

**Vita** **109**

# List of Tables

2.1 DH parameters and joint limits. . . . . 28

3.1 Summary of the data sets used in the experiments. . . . . 52

3.2 Specification of the data sets used in experiments. . . . . 53

3.3 Projection images obtained with central-slice theorem. . . . . 55

4.1 Notations . . . . . 67

4.2 Specifications of synthetic and real CT data sets. . . . . 79

# List of Figures


1.1 Exemplar application in image-guided surgery. . . . . 2
1.2 Example of a skin injury from radiation exposure. . . . . 3
1.3 Topics in the dissertation. . . . . 5

2.1 Illustration of the VRC configuration. . . . . 16
2.2 Spinal phantom with IR markers in the experiment. . . . . 18
2.3 Reconstruction of appearance model. . . . . 20
2.4 Calibration between appearance model and pre-operative CT data. . 21
2.5 Transformation of pre-operative CT data. . . . . 22
2.6 Definition of reference pose. . . . . 23
2.7 Illustration of two modes for determining the optimal C-arm pose. . . 24
2.8 Ellipses fitting of the orbital movement. . . . . 27
2.9 Kinematic chain of C-arm model with a moving basement. . . . . 27
2.10 IR markers on the source of mobile C-Arm. . . . . 29
2.11 Experimental setup. . . . . 31
2.12 Appearance model. . . . . 32
2.13 Accuracy of reconstruction. . . . . 33
2.14 Translation mode for determining the optimal C-arm pose. . . . . 34
2.15 Rotation mode for determining the optimal C-arm pose. . . . . 34
2.16 Precision of mobile C-arm positioning. . . . . 35

3.1 Configuration of tracking based initialization. . . . . 43

3.2 Illustration of the tracking-based initialization. . . . . 44
3.3 Multi-resolution estimation in Fourier domain. . . . . 46
3.4 Wavelet pyramid of the image in golden standard data set Pawiro et al. (2011). . . . . 48
3.5 Illustration of central-slice theorem in 3D. . . . . 48
3.6 Geometry of AP and LAT views. . . . . 49
3.7 Rotation estimation based on multi-resolution strategy using spinal phantom. (a) and (b) represent the mTRE and running time without using multi-resolution strategy, respectively. (c) and (d) are the corresponding performances using multi-resolution strategy. . . . . 56
3.8 Frequency estimation using golden standard data set. . . . . 57
3.9 Results of the complete initialization test. . . . . 58
3.10 Results of 2D/3D intensity-based registration. . . . . 59

4.1 Dimensionality reduction for transfer function design. . . . . 66
4.2 Illustration of dictionary learning. . . . . 68
4.3 Design of multi-dimensional transfer function. . . . . 69
4.4 Illustration of geometry distance in three dimensions. . . . . 70
4.5 Example of neighbor selection using geometry distance. . . . . 73
4.6 Results of conventional methods. . . . . 80
4.7 Results of GCSE. . . . . 80
4.8 Topologies of input data and dictionary in 3D. The blue stars are input data, and the learnt dictionary is indicated by the red stars. . . . . 81
4.9 Residual errors using different configurations on Toroidal Helix data set. 82
4.10 Comparison between sparse embedding and proposed method. . . . . 83

4.11 Rendered images using different transfer functions. (a) is the result using the transfer function based on the intensity and gradient information. (b) shows the rendered image based on the features obtained by GCSE. A, B and C indicate the background/enamel, pulp/dentin, background/dentin boundaries respectively. . . . . 84
4.12 Rendered image of spinal phantom based on intensity and gradient. . 85
4.13 Rendered image of spinal phantom based on proposed method. . . . . 86

# Nomenclature

| | |
|---|---|
| $\alpha_i$ | Coefficient vector in $\mathbb{R}^{k\times 1}$ of $i$th data over $D$ |
| $\lambda$ | Regularization parameter for the sparsity of $\alpha$ |
| $\lambda_G$ | Regularization parameter for the geometrical constraint |
| $\| \bullet \|_F$ | Frobenius norm |
| $\theta$ | Coefficient matrix in $\mathbb{R}^{k\times n}$ over $D$ |
| $D$ | Dictionary in sparse coding in $\mathbb{R}^{m\times k}$ |
| $d$ | The dimension of output space |
| $d_i$ | The $i$th atom in $D$ |
| $G(C)$ | Gramian matrix |
| $K$ | Number of neighbors in the neighbor graph |
| $k$ | The size of dictionary |
| $KK$ | Number of neighbors in the neighbor graph |
| $m$ | The dimension of feature vectors |
| $n$ | The number of voxels in the volume |
| $P$ | Mapping in $\mathbb{R}^{d\times m}$ from input to output space |

| | |
|---|---|
| $Vol(C)$ | Volume of the parallelotope spanned by the clique $C$ |
| $X$ | Features as input data in $\mathbb{R}^{m \times n}$ |
| $Y$ | Mapping features in $\mathbb{R}^{d \times n}$ |

# Chapter 1

# Introduction

## 1.1 Mobile C-arm Positioning in Image-Guided Surgery

In the operating theatre image-guided surgery (IGS) helps the surgeon precisely identify and target the surgical site. It revolutionizes medicine by making a small incision in the patient's anatomy, inserting surgical tools through the incision into the patient's anatomy and using visual aids to guide surgical procedures Maybody et al. (2013). These are performed with minimal invasiveness and enhanced clinical outcomes. Compared to traditional open surgery, IGS reduces scarring, hospital stay duration, blood loss, postoperative pain, recovery time and unnecessary muscle cuts Cleary and Peters (2010).

As shown in Fig. 1.1, an image-guided surgery commonly performs based on a navigation system in which the pre-operative and intra-operative images are involved to present the progress of operation to the surgeon. The pre-operative data are acquired from computed tomography (CT) or/and magnetic resonance imaging (MRI) scan for diagnosis and surgical planning, while X-ray images are usually taken as the intra-operative data to align the pre-operative data with the patient on the clinical table, confirm the surgical outcomes and determine the following surgical decision.

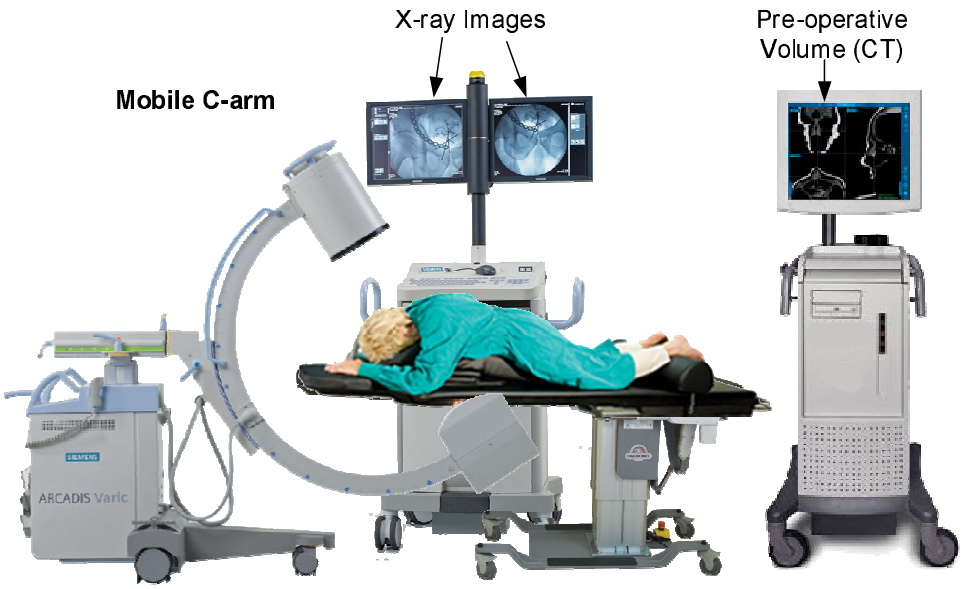



**Figure 1.1:** Exemplar application in image-guided surgery.

The mobile C-arm imaging system is one of the most essential tools for X-ray imaging in the congested operating room due to its compact and versatile structure. During an intervention, the mobile C-arm is activated at distinct points of the procedure to provide the information of current anatomy around the region of interest.

Currently, there are a few commercial intra-operative navigation systems, such as Da Vinci surgical system*, VectorVision† and StealthStation‡, which are capable of registering the treatment plan to the patient and visualizing the volume (i.e., patient's anatomy, physiology and pathology) with respect to the pose of surgical instruments. However, these systems don't take into account the radiation exposure to the patient and surgeons, especially when the X-ray image is taken using the mobile C-arm. In practice, several X-ray images have to be captured repeatedly at a specific surgical stage, because the previous images are blurred or the projection views are not desired. In such cases, the patient and surgeons may be exposed to the unnecessary radiation during the intervention, which can increase the risk of some postoperative

*Intuitive Surgical, Inc., USA. http://www.intuitivesurgical.com/products/davinci_surgical_system/
†BrainLAB, Inc., Germany. http://www.brainlab.com/
‡Medtronic, Inc., USA. http://www.medtronic.com

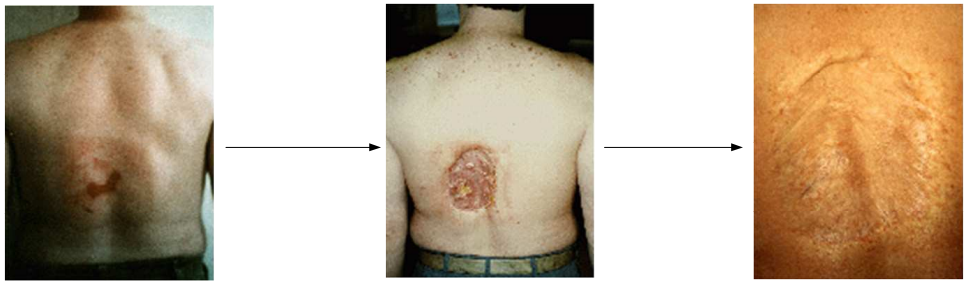


**Figure 1.2:** Example of a skin injury from radiation exposure.

complications Bindal et al. (2008); Synowitz and Kiwit (2006); Stabile et al. (2012), such as skin burns, nausea, cancer and so forth. Fig. 1.2 shows an example of a skin injury from the excessive radiation exposure. The skin turns red about one month after the procedure, and progressive necrosis continues over the following months. A skin graft was required for the injury eventually. Therefore, it is a critical issue to minimize the radiation exposure by positioning the mobile C-arm to a desired pose, and achieve one-time X-ray imaging, which can also greatly reduce the manipulation of C-arm and shorten the total procedure time.

Besides, the positioning of mobile C-arms during an intervention is time consuming and requires considerable skills for the radiation worker due to the complex kinematic constraints of a C-arm and limitation of congested operating room Binder et al. (2006). Particularly in the image-guided surgery, it is difficult to get best viewing projections in regard to the hidden anatomy. Because there is no context with respect to the patient's skin to help surgeons to locate the position of region of interest from the outside view. The radiation worker has to rely on some guess work according to the experience. To some extent, the proficiency level determines the times of X-ray imaging. Besides, it will elongate the learning curve for the novices.

Virtual reality (VR) is a simulated environment where multi-modality data are registered into a common space. It provides more information in regard to the region of interest. In the past decade, it has been employed in surgical training and education, and have played an important role in the training for surgical procedures

Albani and Lee (2007). Since it lacks the information with respect to the patient's outside during the positioning of mobile C-arms, the concept of VR could be applied to align the inside and outside of patient. The radiation worker would have an X-ray vision system that allowed them to view the hidden anatomy and context together, and locate the pathology from the outside view.

As a vital link between the surgeons and multi-modality data in the image-guided surgery, the quality of visualization is also important during the positioning of mobile C-arms. Particularly, a high-quality visualization of pre-operative volume would help the surgeon identify the desired region in the patient's anatomy, and determine the optimal pose of C-arm. Direct volume rendering (DVR) is commonly employed using the transfer function (TF) to emphasize the different structures of the anatomy. Multi-dimensional TFs are able to differentiate the details accurately, while the one-channel TF based only on the intensity in the volume can not discriminate two regions with similar intensity properly. With the help of high-quality visualization of patient's anatomy and context provided by VR, the mobile C-arm could be positioned accurately and fast in the clinical application, and this brings various benefits to the surgeons and patient.

## 1.2 Research Goals

This research aims to develop a VR aided positioning system for mobile C-arms in the image-guided surgery. It is capable of minimizing the radiation exposure to the patient and surgeons, reducing the positioning time, and enhancing the capability of visualization. When the pre-operative volume and patient's surface model are aligned, the positioning system is to determine a reference pose from the outside view. Besides, high-quality visualization enables a better guidance to locate the pathology in the anatomy. There are three goals in total discussed in this dissertation, as shown in Fig. 1.3.

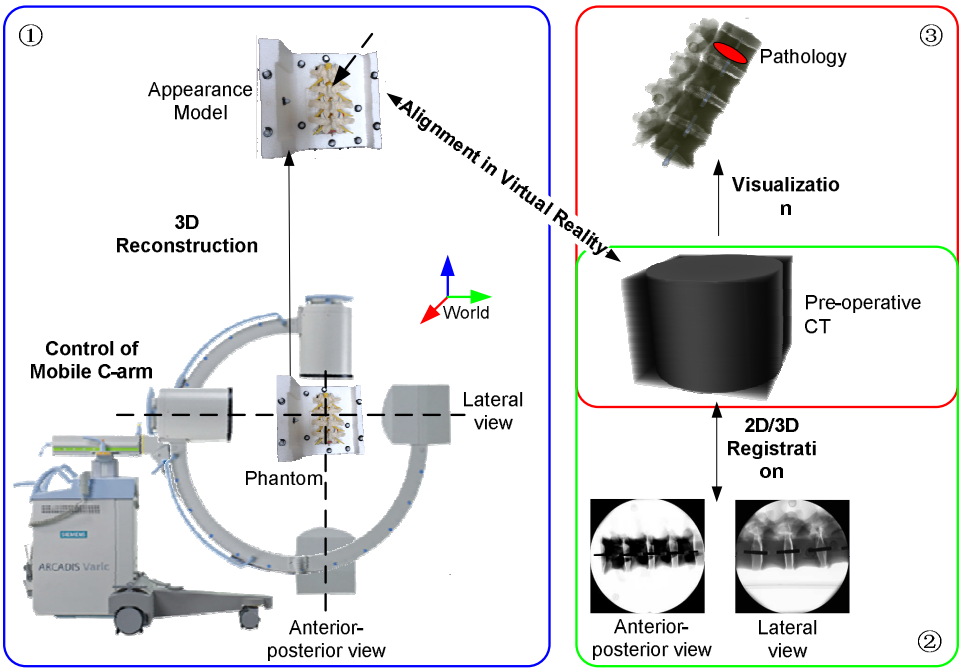



**Figure 1.3:** Topics in the dissertation.

The first goal is to develop a framework for mobile C-arm positioning based on the concept of VR. One key technique is the alignment of pre-operative volume and patient's surface model. The former can be obtained by CT or MRI scan, and the later is usually reconstructed based on a sequence of images from one or more cameras. The fundamental requirements for the surface model reconstruction are: 1) the setup does not occupy the working space of surgeons during the operation; 2) the patient reconstruction can be integrated into the clinical workflow seamlessly. To obtain the physical model of the patient, the positions of camera(s) have to be tracked. In this research, *Vicon motion tracking system* [§] is employed by attaching several infrared reflective markers on the camera. The final patient model is transformed to the coordinate system of Vicon. To register the patient model and pre-operative volume in a common coordinate system (Vicon's frame), intensity-based 2D/3D registration technique is employed. The position of X-ray source is also tracked by Vicon, so the

[§] http://www.vicon.com

resultant transformation is applied in the Vicon's coordinate. Once the alignment is finished, we can obtain a reference pose of mobile C-arm in the Vicon's frame. To facilitate the manipulation, the mobile C-arm is modeled as a mobile manipulator consisting of a mobile base and a robot manipulator with five degrees of freedom. The movements of each joint may be tracked by a visual tracker.

The second goal is to develop a robust and fast initialization method for intensity-based 2D/3D registration to address the issue in the transformation of pre-operative volume. Although the existing intensity-based registration methods are implemented with the high accuracy, a primary problem is the small capture range, which is defined as the 95% successful rate in the order of 2 mm mean target registration error (mTRE). Currently, the field of intensity-based method is dominated by digitally reconstructed radiograph (DRR) based method. Another issue is the computational load caused by the DRR generation. The trial in the frequency domain is explored to avoid this time-consuming procedure.

The third goal is to provide a guidance for the mobile C-arm positioning to locate the region of interest in the anatomy by the high-quality visualization. As aforementioned, DVR is commonly used for 3D medical visualization, and multi-dimensional transfer function is employed to map multiple features to different color and opacity properties. Feature vectors might consist of gradient magnitude, directional second derivative, statistical measures and so on. In this way, more accurate perception can be provided for surgeons, as well as better outcome is enabled in further processing (e.g. data classification and object recognition).

## 1.3 Research Challenges

### 1.3.1 Mobile C-arm positioning

During the C-arm positioning, It is a great challenge to properly align the mobile C-arm to the hidden region of interest (ROI), especially for corpulent bodies. Currently,

the positioning of mobile C-arm extensively depends on the surgeon's experience, as the ROI is in the hidden anatomy, and there is no appearance information as the reference. Surgeons can easily identify the ROI in the CT data, but it is difficult to figure out the pose of mobile C-arm relative to the appearance of patient in image-guided surgery. Although the surgeon makes several marks on the patient's skin for the self-explaining using pre-operative CT data, X-ray images have to be captured repeatedly for the image quality and projection view in practice. This increases the additional radiation exposure to the patient and surgeons, and prolongs the procedure of the whole operation.

Additionally, it is difficult to operate the mobile C-arm manually due to the complicated rotational and linear movements of C-arm. To facilitate the operation of mobile C-arm, it is usually modeled as a manipulator with five or six degrees of freedom (DoFs) in literatures, e.g., robotized C-arm system with 5 DoFs Binder et al. (2006); Matthaus et al. (2007), and 6-DoF C-arm system integrating the operating table Wang et al. (2012b). However, considering the limit of each DoF, mobile C-arm with current modeling cannot reach an arbitrary pose in 3D Cartesian space. And the determination of the target pose of mobile C-arm is not discussed and defined explicitly in literatures.

### 1.3.2 Initialization of 2D/3D registration

For the initialization of intensity-based 2D/3D registration, in order to provide a large capture range, Van der Bom *et al.* employed a coarse registration method in Fourier domain Khalifa et al. (2011); Jerbi et al. (2013) as a robust initialization method for intensity-based registration using the central-slice theorem De Castro and Morandi (1987) and phase correlation Reddy and Chatterji (1996) to estimate the rotational pose of volume and translational offsets, respectively. In the remainder of this paper, this technique is referred to as the frequency estimation. It does not require the DRR generation, and the output provides a set of parameters around the final

optimum, that is, less iterations are required to maximize the similarity measurement in intensity-based registration. However, the initial parameters of rotation estimation have to be set manually. Also according to the reported result in van der Bom et al. (2010), the proposed method took 7.5 min per initialization, and the successful rate of proposed method followed by intensity-based registration was only 68.6%, although it has been improved from 28.6%.

In order to develop a more robust and faster initialization method in frequency domain in an automatic manner for the practical use of 2D/3D intensity-based registration, it is a great challenge to estimate the relative pose of intra-operative 2D images due to the instability of mobile C-arm commonly used in the scenario of IGS. Additionally, the initialization of frequency estimation becomes more important to improve the robustness, and faster estimation is also required to reduce the processing time of whole registration procedure.

### 1.3.3 Visualization

Multi-dimensional transfer function is used for medical visualization. Although more details may be emphasized, as the dimensionality of TF domain increases, the direct interaction with the transfer function becomes impractical. In this case, we have to find a low-dimensional representation of high-dimensional features to enable the effective interaction and preserve enough information of features. One solution to this problem is dimensionality reduction (DR) which is the transformation from the high-dimensional data to a representation in the reduced dimension without significant information loss Van der Maaten et al. (2009). Another advantage of dimensionality reduction is that the redundant information can be filtered out simultaneously. Therefore, an approach of dimensionality reduction is employed to facilitate the design of transfer function in this paper. Conventional approaches of dimensionality reduction in the machine learning field are primarily divided into two categories: linear and non-linear methods. Linear dimensionality reduction is

characterized by the linear mapping, such as principal component analysis (PCA) Jolliffe (2005), classical multidimensional scaling (MDS) Wang (2011), independent component analysis (ICA) Hyvärinen and Oja (2000), and projection pursuit (PP) Jimenez and Landgrebe (1996). The nonlinear methods, including kernel principal component analysis (KPCA) Hoffmann (2007), Isomap Tenenbaum et al. (2000), locally linear embedding (LLE) Roweis and Saul (2000) and Laplacian eigenmaps (LE) Belkin and Niyogi (2001), perform better by taking into account the nonlinearity of the original data. However, both methods incur the problem of heavy computational load on the large data set. For example, assume the dimension of TF domain is 5, and a CT volume with the size of $256 \times 256 \times 128$ is given, the input to dimensionality reduction methods is in $\mathbb{R}^{5\times 8388608}$. It is extremely difficult to deal with in a batch way. Nevertheless, the fast implementation of transfer function design is required in both pre- (diagnosis and surgical planning) and intra-operation.

Recently, the rapid development in the filed of sparse representation (SR) paves another way to the dimensionality reduction. This method is achieved based on the assumption that the signals are compressible, and have a sparse representation in a basis set (also a.k.a dictionary). It has been proven practical in many applications Agarwal et al. (2004); Starck et al. (2005); Elad (2010). In Zhang et al. (2010), given a dictionary, a sparse dimensionality reduction scheme was presented as a regressive problem, and the optimal mapping was found by minimizing the cost function of sparse coding. A similar concept was presented in Nguyen et al. (2012), Nguyen *et al.* proposed a sparse embedding (SE) framework by combing dimensionality reduction and sparse learning together. It enables the sparse representation (sparse coding and dictionary learning) in the reduced dimension with lower computational load. However, current SR based methods are still time-consuming as the mapping matrix is optimized together with SR, especially for the medical data with the large size.

To achieve a SR promoting dimensionality reduction method with the capability to process the large data set efficiently for the transfer function design in the direct volume rendering, it is a great challenge to reduce the computational burden, when

mapping matrix is integrated into the SR framework. Another difficulty is to find a optimal dictionary, which can better describe the features extracted from the medical data.

## 1.4 Contributions

The positioning of mobile C-arm is discussed in this dissertation, The main contributions are summarized as follows.

1. A Visual Reality (VR) aided positioning system is proposed. The mobile C-arm is augmented by a single camera, which can reconstruct the patient's surface model using the Structure from Motion (SfM) and Multi-view Stereo (MVS) techniques. The mobile C-arm is modeled as a mobile manipulator, and the movement of each joint is tracked by Vicon motion tracking system. Detailed contributions are highlighted as follows:

   (a) VR aided positioning system is developed to determine a reference pose for the mobile C-arm from the outside view of ROI. One-time imaging reduces the unnecessary radiation to the patient and surgeons, and the probability of complication probability.

   (b) The mobile C-arm is augmented with a single camera, and a well-designed calibration procedure is proposed to register the mobile C-arm and single camera. The calibration enables the alignment of pre-operative volume and patient's model in Vicon's coordinate system.

   (c) The mobile C-arm is modeled as a redundant robot by adding three more degrees of freedom onto the base. It significantly increases the flexibility and mobility in the congested operating room.

2. A two-level initialization method is proposed including tracking based initialization (Level I) and finer estimation based on multi-resolution strategy in

frequency domain (Level II). It allows the fast implementation and improves the robustness of estimation by starting registration from the coarse level. Detailed contributions are summarized as follows:

(a) Automatic estimation of 2D/3D intensity-based registration is proposed based on tracking. Surgeons just take the intra-operative images at the distinct time, the following tasks can be done automatically to link the pre- and intra-operative data.

(b) More robust and faster frequency estimation for intensity-based registration is achieved based on the wavelet pyramid in the Fourier domain.

(c) The proposed method not only addresses the problem in the mobile C-arm positioning, but also can be apply to similar case and other registration methods.

3. A novel framework of Geometry Constrained Sparse Embedding (GCSE) is proposed for the design of transfer function in DVR. Since GCSE is a SR promoting method, the problem is reduced to the optimization of dictionary. To this end, the intrinsically geometrical structure based on a geometry distance are taken into account. Detailed contributions are summarized as follows:

   (a) a novel framework based on sparse representation for dimensionality reduction is proposed. The geometrical structure of features are preserved. To the best of our knowledge, it is the first time to introduce the sparse representation into the design of transfer function in direct volume rendering;

   (b) a geometry distance is proposed to select neighbors with geometrical relationships in the high-dimensional space, and a neighbor graph is constructed based on the geometry distance to find a optimal dictionary;

(c) the proposed method separate the computing of mapping matrix from SR, it improves the capability to deal with the large data set in an efficient way.

## 1.5 Dissertation Outline

This dissertation consists of three related papers which are arranged in the order as research goals summarized in Section 1.2. It is organized as follows:

Chapter 2 presents the framework of VR aided positioning system for mobile C-arm, including the alignment of inside and outside of patient, kinematic model and control of mobile C-arm.

Chapter 3 introduces an initialization scheme to speed up the intensity-based 2D/3D registration and improve the accuracy of registration. A hardware-software integrated approach is proposed, including the tracking based initialization and multi-resolution estimation in frequency domain.

Chapter 4 introduces a novel way to design a transfer function in the direct volume rendering. Dimensionality reduction method is employed to obtain a new representation of feature vectors in a reduced dimension, while preserving most information of features.

Finally, Chapter 5 concludes the dissertation and gives description of future works.

# Chapter 2

# Visual Reality Aided Positioning of Mobile C-arms

# Abstract

For the image-guided surgery, the positioning of mobile C-arms is a key technique to take X-ray images in a desired pose for the confirmation of current surgical outcome. Unfortunately, surgeons and patient often suffer the unnecessary radiation exposure due to the repeated imaging when the X-ray image is of poor quality, or not captured at a good projection view. Besides, the mobile C-arm positioning currently need to be performed in a limited working space with the high requirement of time and skills for the radiation worker. In this paper, an Visual Reality (VR) aided positioning method for the mobile C-arm is proposed by the alignment of 3D surface model of region of interest and pre-operative anatomy, so that a reference pose of the mobile C-arm with respect to the inside anatomy can be figured out from outside view. It allows a one-time imaging from the outside view to greatly reduce the additional radiation exposure. Surgeon's hands relative to the X-ray cone are also taken into account. To control the mobile C-arm to the desired pose, the mobile C-arm is modeled as a robotic arm with a movable base. Three more degrees of freedom assigned to the base of C-arm, that increases the mobility and flexibility in the congested operating room. Experiments were conducted to evaluate the accuracy of appearance model and precision of mobile C-arm positioning. The appearance model was reconstructed with the average error of 2.16 mm. One-time imaging of mobile C-arm was achieved, and new modeling of mobile C-arm with 8 DoFs enlarges the working space in the congested operating room.

## 2.1 Introduction

Image-guided surgery (IGS) is a minimally invasive procedure by making small incisions on patient's body Cleary and Peters (2010). In IGS, pre-operative CT data and intra-operative X-ray images are widely used for the intervention guidance. Pre-operative data is used for diagnosis and surgical planing, and X-ray images is captured to verify the surgical outcome. The mobile C-arm is one of the most essential tools for X-ray imaging in the congested operating room due to its compact and versatile structure. During the operation, X-ray images can not be captured continuously due to the radiation exposure. To reduce it, the mobile C-arm need to be positioned in a proper pose at distinct points of operation. Accurate positioning technique of mobile C-arm is required in surgical applications, e.g., percutaneous needle procedures and fracture fixation with locking screws and nails. In such cases, the C-arm needs to be piloted to the needle progression view and entry point view to confirm the current pose of the needle Leschka et al. (2012), or over the nail so that each screw hole appears as a perfect circle Bott et al. (2011). Therefore, mobile C-arm positioning is a routine surgical task, and has been become more important for the intra-operative imaging techniques Wang et al. (2012b).

The objective of this paper is to determine a reference pose relative to the patient's appearance, and achieve one-time imaging using mobile C-arm. To this end, a Virtual Reality (VR) aided mobile C-arm positioning method is proposed. Fig. 3.1 illustrates the configuration of proposed method. In VR, an appearance model of patient and pre-operative CT data are aligned, so that a reference pose is determined in regard to the appearance. The appearance model is reconstructed by a single camera mounted on the mobile C-arm. To facilitate the control of mobile C-arm to the reference pose, the mobile C-arm is modeled as a mobile manipulator. The movements of mobile C-arm and camera are tracked by Vicon motion tracking system referred to as Vicon for simplicity in the remainder of this paper.

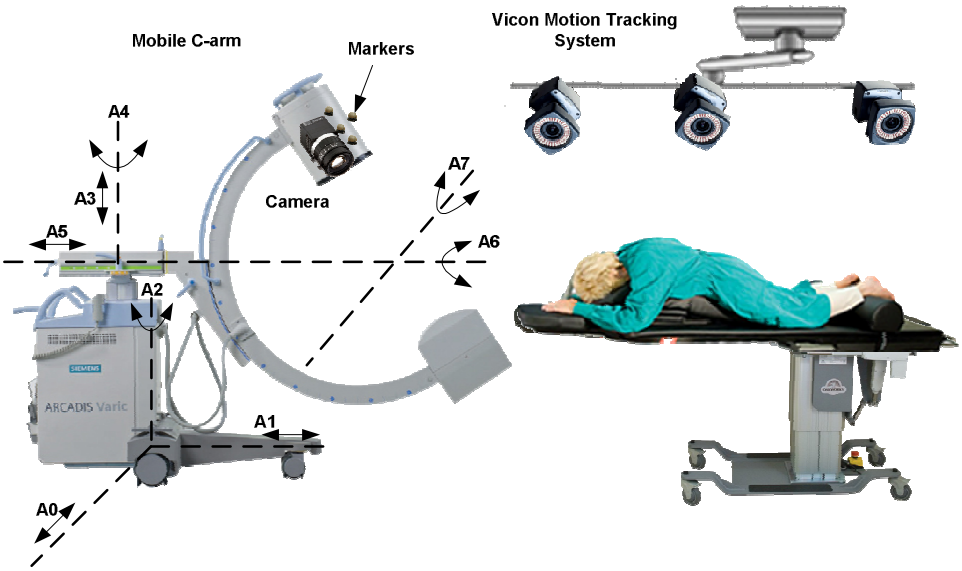



**Figure 2.1:** Illustration of the VRC configuration.

## 2.2 Materials

To determine a reference pose relative to the appearance of patient, the mobile C-arm is augmented by mounting a single camera on its gantry next to the X-ray source. An appearance model of patient can be reconstructed from images recorded by the camera. Then the appearance model and pre-operative CT data are aligned in virtual reality to find the reference pose, since the region of interest may be identified in pre-operative data. To achieve the reference pose, the mobile C-arm is controlled as a mobile manipulator consisting of a mobile base and a robot manipulator with five DoFs (A3-A7). The mobile base has additional three DoFs, including both translations (A0 and A1) in the transverse plane, and rotation against the gravity direction (A2).

In this paper, a mobile C-arm, single camera, Vicon motion tracking system and spinal phantom are used. The camera is mounted on mobile C-arm to capture the appearance information of patient, which acts as a reference for mobile C-arm positioning. Because the mobile C-arm is modeled with 8 DoFs to facilitate the

operation of C-arm, and it is not motorized yet, Vicon is employed to track the movements of joints on mobile C-arm. A spinal phantom was chosen in experiments to verify the performance of proposed method.

### 2.2.1 Camera Augmented Mobile C-arm

ARCADIS Varic (Siemens, Germany) is used as a advanced multi-purpose mobile C-arm. It enables $1K^2$ X-ray image. The mobile C-arm is usually modeled as a manipulator with 5 degrees of freedom, including vertical, swivel, horizontal, angular and orbital movement of the image intensifier. Detailed illustration can be found in Chaganti et al. (2009). In order to track the mobile C-arm, Vicon is employed, and several infrared (IR) reflective markers are attached on the C-arm. These markers can be used to determine the increment of each joint, the position of C-arm base and the pose of X-ray source. The tracking system will be introduced later.

The mobile C-arm is augmented with Guppy Pro F-201C from AVT Inc., Germany. This camera provides high-quality images with the resolution of 2 Megapixel ($1624 \times 1234$). The camera was also tracked by Vicon to reconstruct an appearance model with the physical size.

### 2.2.2 Vicon motion tracking system

Vicon motion tracking system (Vicon Inc., UK) consists of Vicon T-Series cameras that are the world's next generation motion capture devices. Each camera provides a 16 megapixels resolution ($4704 \times 3456$) at the frame rate of 120 fps. The tracking system works with IR reflective markers that are different size of small plastic balls covered with reflective tape from 3M Inc., US. When the infrared is projected to the surface of markers, the light with the same wavelength returns to the camera, so that each camera can get the 2D position of markers, then the 3D position can be obtained with sub-milimeter accuracy by the controller of Vicon. Proper size of makers are

chosen in practice, according to the requirement of applications. The diameter of makers used in our experiments is 14 mm.

### 2.2.3 Spinal phantom

The spinal phantom (3B Scientific Inc., Germany) offers a illustration of lumbar vertebrae from L1 to L5 with intervertebral discs. The finest bone structures are accurately depicted. To fix the spine, the model is placed in the pearl cotton, as shown in Fig.2.2. 11 IR reflective markers are attached to facilitate the process of calibration between the Vicon and pre-operative CT data. The CT data is collected by a CT scanner (Aquilion ONE, Toshiba, Japan) with a 320-row detector. The resolution of each slice is $512 \times 512$ pixels, and there are 227 slices in total. The CT volume covers the region of $0.559 \times 0.559$ mm with a 1-mm slice thickness.

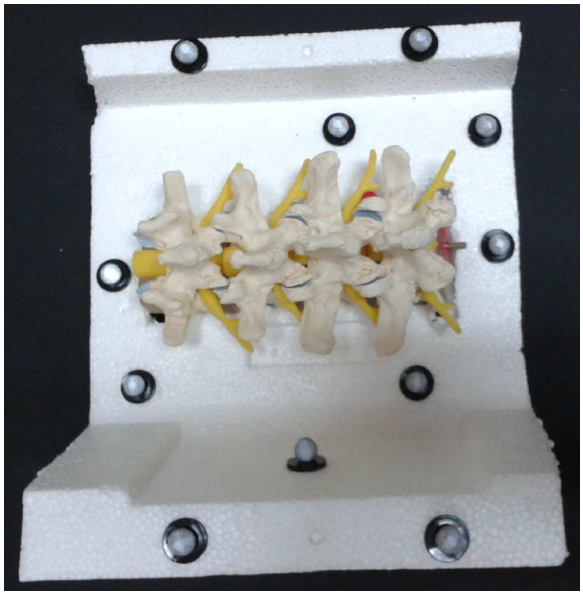


**Figure 2.2:** Spinal phantom with IR markers in the experiment.

## 2.3 VR Aided Determination of Reference Pose

In this section, the reference pose of mobile C-arm relative to the appearance is figured out based in VR. An appearance model is firstly reconstructed using the camera on mobile C-arm. Then the appearance model and pre-operative CT data are aligned in coordinate system of Vicon. Above procedures need to be implemented only once at the beginning of operation. Since the region of interest can be found in CT data, a reference pose in regard to appearance can be determined using an optimal procedure.

### 2.3.1 Reconstruction of Appearance Model

Along with the rotation of mobile C-arm, several images are taken by the camera at different positions, as shown in Fig. 2.3. Orientations of camera are converged to the target. To reconstruct the 3D model, approaches of Structure from Motion (SfM Snavely et al. (2006)) and Multi-view Stereo (MVS Seitz et al. (2006)) are employed in this paper.

Some principles have to be considered during the image collection for SfM-MVS reconstruction. Static scene is required as the reconstruction works by matching the features from different images. Another factor is the illumination condition, strong shadows should be avoided. The reconstruction of patient or ROI totally meets above requirements in the operation room. During the operation, the patient is static due to the anesthesia, and there is no strong shadows when using the surgical shadowless lamp.

In this paper, a sparse point cloud is firstly produced based on a SfM application *Bundler* Snavely et al. (2006). SIFT features are extracted from all images, then the most prominent features are matched in other images. The 3D coordinates can be computed based on the feature pairs. The resulting point cloud only represents the positions of limited amount of features. To enhance the previous model, Clustering Views fro Multi-view Stereo (CMVSFurukawa et al. (2010)) is used for a dense 3D model. It uses the output of *Bundler* to decompose the input images into a set of

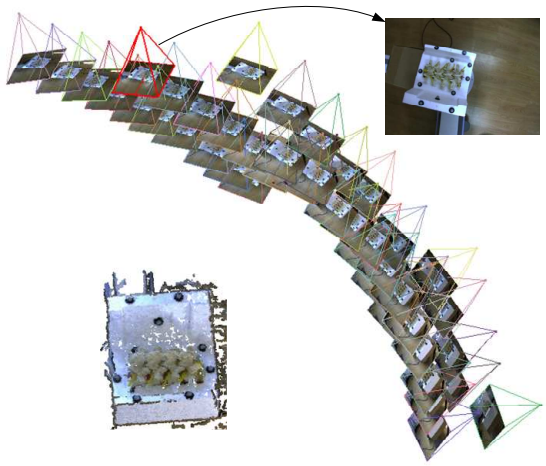


**Figure 2.3:** Reconstruction of appearance model.

clusters. A patch-based MVS algorithm (PMVS Furukawa and Ponce (2010)) can be used to process each cluster independently and in parallel. This process can efficiently filter out the noise from the SfM process, and increase the number of reconstructed points. Nevertheless, the resulting model lacks the scale information compared with the physical object. To this end, the positions of camera are tracked by Vicon during the image collection.

### 2.3.2 Calibration of Appearance Model and Pre-operative CT

In this paper, the appearance model and pre-opertaive CT data are calibrated in Vicon's frame, as shown in Fig. 2.4. Since the camera is tracked by Vicon, the appearance model can be transformed to Vicon's frame easily. To transform pre-operative data to a common coordinate system, intensity-based 2D/3D registration is carried out using 2D X-ray images, which are captured at the beginning of operation.

In 2D/3D registration, we build a virtual scene identical to actual setting, including focal point (X-ray source) and image plane, as shown in Fig. 2.5. Applying the transformation to pre-operative CT, we can generate a floating image by casting

rays through CT. The value of each pixel is the summation of intensities along the projection ray. Given the pose of focal point and image plane in Vicon's frame, the transformation of CT data is obtained in Vicon's frame by optimizing the similarity between X-ray image and floating image.

### 2.3.3 Reference Pose Optimization of Mobile C-arm

The reference pose of mobile C-arm relative to the ROI in pre-operative CT is determined firstly. Since the appearance model and CT data are aligned, we can get the reference pose in regard to the appearance. A vector $[x, y, z, \alpha, \beta, \gamma]$ is defined to denote the pose of mobile C-arm, including the position of X-ray source and orientations between central X-ray beam and axes of the coordinate frame of Vicon, as shown in Fig. 2.6. There are infinite solutions for the reference pose. In this paper, the reference pose is optimized by the image quality and radiation exposure. According to as-low-as-reasonably-achievable (ALARA) principle van der Merwe (2012), the patient is placed as close to image intensifier and as far from x-ray tube as possible. Because maximum scatter reflects from the side of the patient that is closest to the x-ray source. The beam is directed through the patient at a perpendicular angle to reduce the radiation dose due to the increased tissue thickness. Additionally, the surgeons sometimes have to operate the instrument to finish the specific surgical tasks

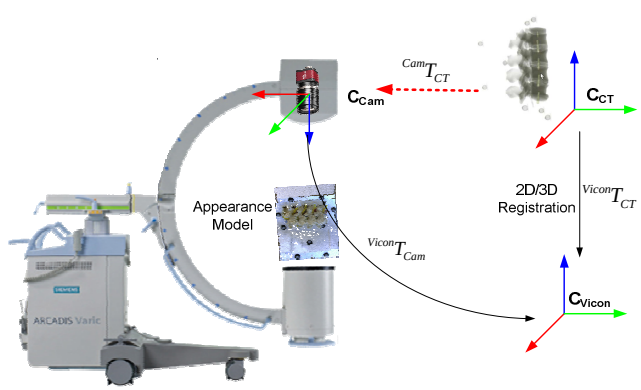



**Figure 2.4:** Calibration between appearance model and pre-operative CT data.

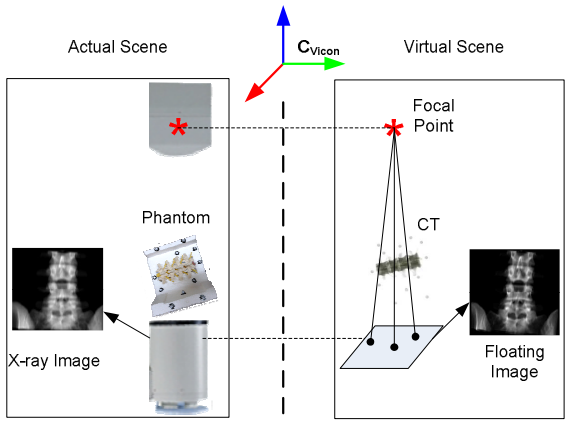



**Figure 2.5:** Transformation of pre-operative CT data.

---

**Algorithm 1** Calculate the optimal pose of X-ray source.

---

**Input:** Specific region, radius of the image intensifier $r$, distance between X-ray source and image intensifier $d$, hand's position $H$, moving range of X-ray source $[near, far]$, maximal angle shift of X-ray source with respect to ROI $shiftMax$

**Methed:**

Calculate the normal across the center of specific region $\overrightarrow{n}$ using VTK.
Calculate the parametric equations $g(x, y, z)$ of circle $O$ according to $r, d$ and $\overrightarrow{n}$.

Model the X-ray beam as a cone formulated by the vertex $A(x0, y0, z0)$ and the directrix $g(x, y, z)$.
Calculate the intersecting points $M, N$.
**if** $\overrightarrow{MH} \cdot \overrightarrow{NH} \leq 0$ at the near point **then**
    Rotation mode is triggered (Fig. 2.7a).
    Minimize min(shiftRot)/max($norm(\overrightarrow{MH})$) to update the axis of X-ray beam.
**else if** $\overrightarrow{MH} \cdot \overrightarrow{NH} \geq 0$ at the far point **then**
    **return**
**else**
    Translation mode is triggered (Fig. 2.7b).
    **if** $norm(\overrightarrow{MH}) > \varepsilon$ **then**
        **return**
    **end if**
**end if**
($\alpha$, $\beta$, $\gamma$) is derived from the axis of X-ray beam, and (x,y,z) is the final position of vertex of cone.

**Output:** Optimal pose of X-ray source **O**(x, y, z, $\alpha$, $\beta$, $\gamma$).

---

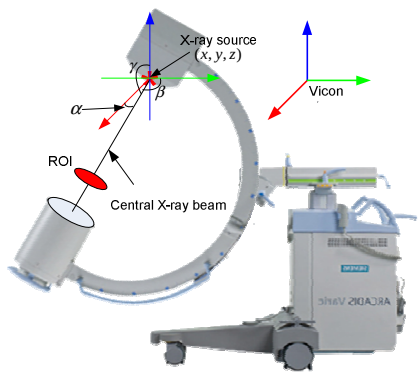


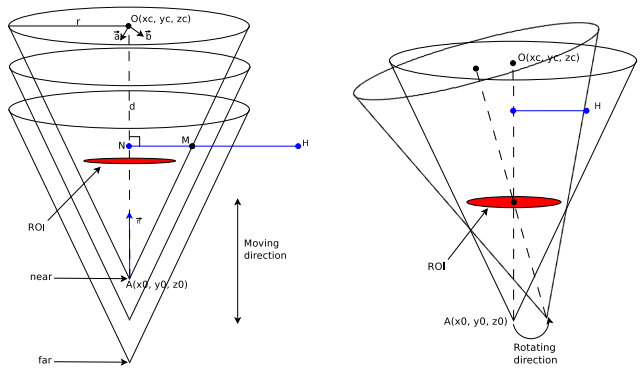

**Figure 2.6:** Definition of reference pose.

near the X-ray beam. This is unavoidable to suffer the radiation exposure. Thus, the positions of surgeons' hands are considered during the determination of the reference pose to enable the hands far from the X-ray cone.

The X-ray beam is modeled as a cone with the convex at the X-ray source, and the directrix which is the outer edge of the imaging plane of the mobile C-arm. After the surgeon specifies the region to be captured, including the position and orientation, the X-ray source of mobile C-arm should be moved to a desired pose $(x, y, z, \alpha, \beta, \gamma)$. The optimal pose of mobile C-arm is determined by considering the image quality, X-ray cone and position of surgeon's hand. The detailed procedure is described in Algorithm 1.

Firstly, the normal across the center of specific ROI $\overrightarrow{n}$ is calculated. To model the cone of the X-ray, the circle $O$ in Fig. 2.7 needs be found, and its parametric equations $g(x, y, z)$ can be denoted as follows.

$$\begin{cases} x(u) &= x_c + r(\overrightarrow{a}_x cosu + \overrightarrow{b}_x sinu) \\ y(u) &= y_c + r(\overrightarrow{a}_y cosu + \overrightarrow{b}_y sinu) \\ z(u) &= z_c + r(\overrightarrow{a}_z cosu + \overrightarrow{b}_z sinu) \\ x_c &= x_0 + d\overrightarrow{n}_x \\ y_c &= x_0 + d\overrightarrow{n}_y \\ z_c &= z_0 + d\overrightarrow{n}_z \end{cases} \tag{2.1}$$




**(a)** Translation mode. **(b)** Rotation mode.

**Figure 2.7:** Illustration of two modes for determining the optimal C-arm pose.

where $(x_c, y_c, z_c)$ and $(x_0, y_0, z_0)$ are the center of circle and the vertex of cone respectively. $\overrightarrow{a}$ and $\overrightarrow{b}$ are two vectors perpendicular to $\overrightarrow{n}$ in the plane of image intensifier. The radius of circle and distance between X-ray source and image intensifier are denoted by $r$ and $d$. To reduce the radiation dose to the surgeon further, we increase the radius of circle corresponding to the image intensifier. Here, we let $r = 150$ mm, $d = 999.8313$ mm, and then the X-ray beam is modeled as a cone formulated by the vertex $A(x0, y0, z0)$ and the directrix $g(x, y, z)$.

$$\begin{cases} x(u,v) &= vx(u) + (1-v)x_0 \\ y(u,v) &= vy(u) + (1-v)y_0 \\ z(u,v) &= vz(u) + (1-v)z_0 \end{cases} \tag{2.2}$$

As a matter of experience, we can obtain the high-quality X-ray image in the range of $[near, far]$, which is the distance from ROI to X-ray source. The optimal reference pose can be found in $[near, far]$ by applying two modes, i.e., translation and rotation mode, as shown in Fig. 2.7. In translation mode, the X-ray source moves along axis of X-ray beam to find the best pose by maximizing the distance from the hand's position $H$ to intersecting point $M$. If there is no candidate in translation mode, the rotation mode is triggered, where the X-ray source moves around a circle with the origin at ROI center and radius measured from position of X-ray source to

ROI center, and at the beginning, the x-ray source is positioned to the $far$ point and the axis of X-ray beam is coincident with the normal of ROI. Subsequently, two modes alternately run to minimize the cost function

$$f(\mathbf{O}) = min(min(shiftRot)/max(norm(\overrightarrow{MH}))) \tag{2.3}$$

where $\mathbf{O}$ is a vector $[x, y, z, \alpha, \beta, \gamma]$, and the angle shift in the rotation mode is denoted by $shiftRot$. Because the CT data and appearance model have been aligned, we can obtain the pose of X-ray source with respect to the appearance. To make sure the projection view is desired, the floating image is generated using the Siddon-Jacob ray-tracing algorithm Jacobs et al. (1998).

## 2.4 Kinematic Modeling and Control of the Mobile C-Arm

To define a unique pose of the image intensifier, it requires six independent parameters resulting in a homogeneous matrix. Obviously, current configuration cannot guarantee an arbitrary target pose to be reached. Moreover, considering the congested operating room, several obstacles need to be avoided to obtain the desired the projection. Therefore, the mobile C-arm is modeled as a redundant robot with three more degrees of freedom assigned to the mobile base, so that the pose of image intensifier is optimized to satisfy the requirements of image quality and radiation doses reduction considering the current position of the surgeon's hands, ROI and cone of the image intensifier. In this paper, ARCADIS Varic introduced by Siemens Medical Solutions is used. Although this specific C-arm is studied in following sections, kinematic principles can be easily applied to other C-arms due to the similar structure.

### 2.4.1 Forward Kinematics of Mobile C-arm

A kinematic chain is modeled for C-arms as a mobile manipulator to define the relation between the joints and the pose of the X-ray source. To model the mobile C-arm accurately, To model the mobile C-arm exactly, coordinate origins have to be identified firstly. Specifically, the origin of orbital joint cannot be found on the body of C-arm, while remaining movements can be determined easily. To this end, multiple infrared reflective markers are amounted on the X-ray tube, as shown in Fig. 2.10, and Vicon is employed to collect the 3D information of markers during the orbital rotation. To simplify the fitting procedure, we place the C-arm along the X-axis to make sure the C-arm is rotation in the X-Z plane, when the X-ray source is directly above the image intensifier. Assume the trajectory is an ellipse, and the direct least squares fitting Halíř and Flusser (1998) is implemented. The ellipse can be represented in conic form, as expressed as follows.

$$\begin{aligned} &0.7081x^2 - 0.0012xy + 0.7062y^2 \\ &+ 695.3x - 1401.3y + 561065.6 = 0 \end{aligned} \tag{2.4}$$

Fig. 2.8 illustrates that the trajectory is an ellipse with 656.4270 $mm$ for major axis and 655.3992 $mm$ for minor axis. It is an approximate circle, so that we can use the center of circle as the origin of orbital joint.

Vicon is used to measure the length of links, especially $a_1$, $a_2$, $a_3$ and $a_4$. $a_1$ denotes the distance between the origins of defined horizontal and angular movement, and $a_2$ is the distance from the origin of angular movement to orbital origin. $a_3$ and $a_4$ are used to define the relation between the orbital joint and X-ray source as the end effector, where $a_3$ represents the distance from origin of orbital joint to the center of X-ray beam, $a_4$ is the distance between the center of X-ray beam and origin of end effector. Following the Denavit-Hartenberg (DH) rules Craig (1989), a set of coordinate systems are assigned to each joint, as shown in Fig. 2.9. The origin of the coordinate frame 0 is assigned on the C-arm base as the world frame.

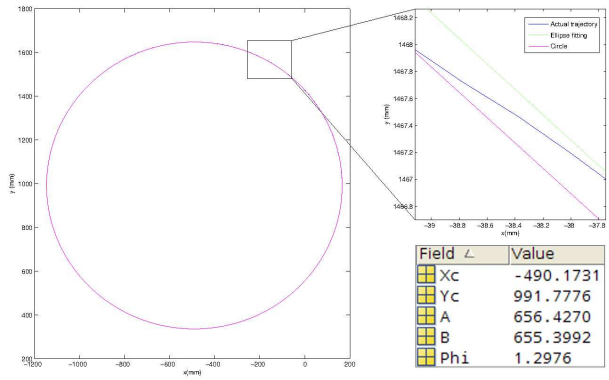



**Figure 2.8:** Ellipses fitting of the orbital movement.

$O_1, O_2, O_3$ are defined as origins of the additional joints. Remaining frames $4 \sim 8$ are same with the standard modeling. $O_e$ is the origin of the X-ray source. Here $a_1 = 1160.8982\ mm, a_2 = 102.9889\ mm$.

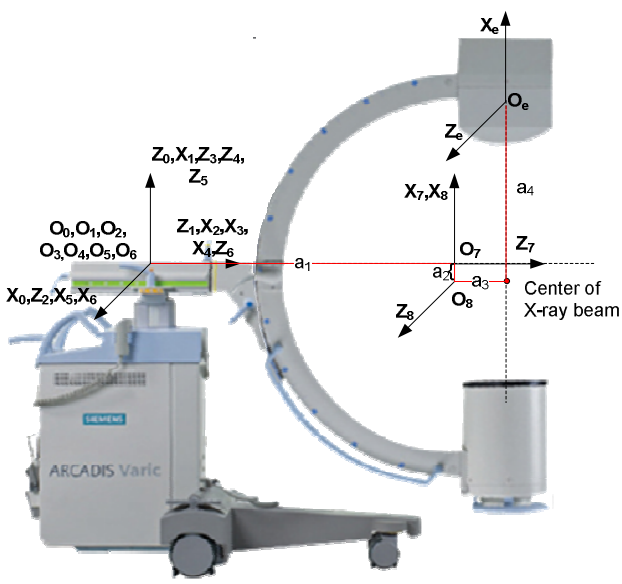



**Figure 2.9:** Kinematic chain of C-arm model with a moving basement.

The corresponding joint parameters and limits are shown in Table 2.1. A vector $q = [d_1, d_2, \theta_3, d_4, \theta_5, d_6, \theta_7, \theta_8]$ is defined to represent the C-arm joint variables. Let ${}^1T_8(q)$ represent the transformation from the orbital joint to the C-arm base, which can be written by

$$^{1}T_8(q) =^{1} T_2 \cdot^{2} T_3 \cdots^{7} T_8 \tag{2.5}$$

where $^{i}T_{i+1} \in R^{4\times4}$ is the transformation from the coordinate system $i+1$ to $i$. It can be derived using the DH parameters in Table 2.1. To define the pose of X-ray source in the world frame and simplify the DH parameters, the coordinate system of X-ray source is defined in accordance with the direction of the orbital joint, so that only translation is considered between them. The origin is at the X-ray source center. Let $^{8}T_e$ be the transformation from the X-ray source to the coordinate system of orbital joint, which can be defined as a constant transformation.

$$^{8}T_e = Trans_y(-a_3) \cdot Trans_x(a_4) \tag{2.6}$$

where $Trans_y(a) \in R^{4\times4}$ represents the translation along $Y$ axis with the increment of $a$. Then, the relation between the X-ray source and world frame can be defined by $^{1}T_e =^{1} T_8 \cdot^{8} T_e$ with respect to the vector $q$. Once the pose of X-ray source is determined, an optimal solution of vector $q$ can be found based on the inverse kinematics method.

**Table 2.1:** DH parameters and joint limits.

| | $\alpha_{i-1}$ (degree) | $a_{i-1}$ (mm) | $d_i$ (mm) | $\theta_i$ (degree) | Limits |
|---|---|---|---|---|---|
| 1 | $-90$ | 0 | $d_1$ | $-90$ | – |
| 2 | $-90$ | 0 | $d_2$ | $-90$ | – |
| 3 | 0 | 0 | 0 | $\theta_3$ | – |
| 4 | 0 | 0 | $d_4$ | 0 | $-200 \sim 250$ mm |
| 5 | 0 | 0 | 0 | $-90+\theta_5$ | $-12.5° \sim 12.5°$ |
| 6 | $-90$ | 0 | $d_6$ | 0 | $-100 \sim 100$ mm |
| 7 | 0 | 0 | $a_2$ | $-90+\theta_7$ | $-180° \sim 180°$ |
| 8 | $-90$ | $-a_1$ | 0 | $\theta_8$ | $-90° \sim 45°$ |

**(a)** Left side. **(b)** Right side.

**Figure 2.10:** IR markers on the source of mobile C-Arm.

### 2.4.2 Inverse Kinematics with Base Movement Control

Given a reference pose of the mobile C-arm, the joint variables can be obtained by solving the inverse kinematics problem. In this section, two cases are considered: (1) the base of mobile C-arm is fixed; (2) the base of mobile C-arm is movable. In this paper, the mobile C-arm is not motorized yet, surgeon operates it directly under the guide of Vicon. Several IR markers tracked by Vicon are attached on the C-arm body to monitor the source pose, base movement and joint increments. Fig. 2.10 shows the markers on the X-ray source to determine the coordinate system of the end-effector. The middle point of the line $S1S2$ is the origin of frame. Z axis points to point $S2$ along the line $S1S2$, and X-axis parallels the vector $\overrightarrow{S3S5}$. During the operation of C-arm, the pose of C-arm is tracked in real time. The current pose provides a feedback to estimate the velocity relative to the target, and then computer the increment of each joint next time. This close-loop procedure is performed until the X-ray source reach the desired pose.

#### Base of mobile C-arm is fixed

The mobile C-arm becomes a manipulator with 5 DoFs, when three DoFs on the base are lost. In the 3D Cartesian space, given a reference pose, an analytic solution can be obtained for each joint. Then we can adjust each joint of mobile C-arm to

achieve the reference pose using Vicon for guidance. In this case, please note that the working space of mobile C-arm is limited around the surgical table.

### Base of mobile C-arm is movable

To enlarge the working space, the mobile C-arm is modeled as a redundant manipulator with a movable base. In this paper. the Closed-Loop Inverse Kinematics (CLIK) algorithm Siciliano (1990) is used, and the general solution of inverse kinematic is employed based on pseudo-inverse Jacobian Chiaverini et al. (1991) to minimize the norm joint velocities and avoid the kinematic singularities. As a redundant manipulator for the mobile C-arm, the redundancy is solved using the Gradient Projection Method (GPM) Liegeois (1977). Thus, the CLIK algorithm with the Proportional-Derivative (PD) feedback can be expressed as follows.

$$\begin{cases} \dot{q} &= J^{\dagger}(\dot{x_d} + K_p(x_d - x)) + (I - J^{\dagger}(q))\dot{q_0} \\ \dot{q_0} &= (\dfrac{\partial(h(q))}{\partial(q)})^T \end{cases} \tag{2.7}$$

where $x_d$ and $x$ are the desired pose and current pose of X-ray source respectively. $K_p$ is a symmetric positive matrix to guarantees the error converges to zero. $h(q)$ is a scalar cost function, which can be optimized to solve the redundancy based on GPM. In addition, the joint limit avoidance is considered for the local optimization, and defined by

$$\Phi(q) = max(\frac{q_i - \overline{q}_i}{q_{iM} - q_{im}}) \tag{2.8}$$

Once the joint variables are determined, the base of C-arm is moved first, then the remaining joints are operated, and eventually the X-ray source reaches the target.

## 2.5 Experimental Results and Discussions

The experimental system was set up in Shenyang Institute of Automation, Chinese Academy of Sciences, China. The mobile C-arm was redesigned using several IR reflective markers on the X-ray source, base, swivel link and corresponding joints, so that the whole C-arm can be tracked using Vicon. Note that a part of markers are used in common during the tracking, e.g., vertical, swivel, horizontal movements. Vicon is fixed on the ceiling of the room in order to track the movement of mobile C-arm, as illustrated in Fig. 2.11. The orientation of Vicon cameras must be adjusted to make sure each IR marker can be seen at any time by three cameras at least for the stable measurement. During the adjustment of tracking system, sometimes the marker on the X-ray source is out of the filed of view when the mobile C-arm is rotated along the orbital joint, especially. The spinal phantom attached makers is placed on the table.

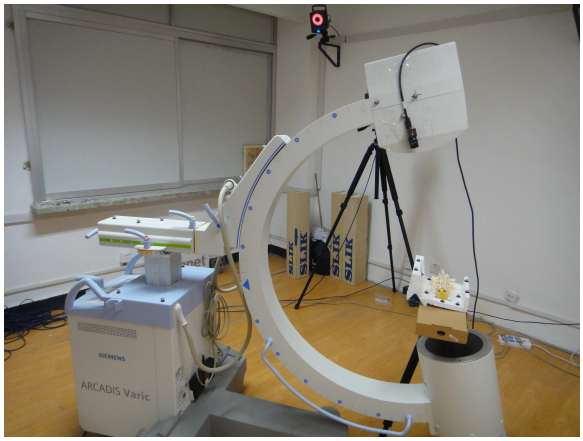



**Figure 2.11:** Experimental setup.

In this section, the accuracy of appearance model reconstructed by the camera was estimated firstly. A simulation was performed to present the procedure of reference pose determination. To demonstrate the accuracy of virtual reality aided mobile C-arm positioning, the similarity between the floating image and X-ray image was measured. All computations in the experiments were performed on the powerful

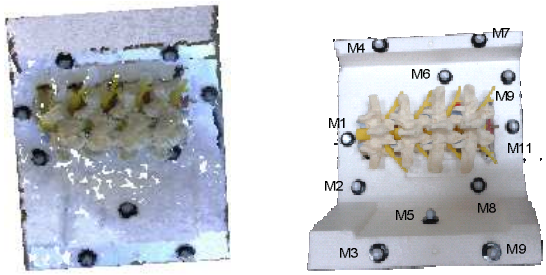
Lenovo ThinkStation D20, which is assembled with Intel Xeon Processor X5650, 12G memory and NVIDIA Quadro 600 graphics card.

### 2.5.1 Accuracy of Appearance Model

Fig. 2.12a shows the result of reconstruction based on multi-view stereo technique. Markers are labeled with $Mi, i = 1, 2 \cdots 11$, as shown in Fig. 2.12b. To measure the accuracy of appearance model, the relative distances from $M1$ to the rest of markers were calculated. Markers were tracked by Vicon as the ground truth. Fig. 2.13 shows the accuracy of appearance model. The average error of reconstruction is 2.16 mm. A accurate model was reconstructed using a single camera in this experiment.

### 2.5.2 Simulation of reference pose determination

The simulation of reference pose estimation is performed using Matlab. In the simulation, besides the ALARA principle, the minimal distance between the hand ($H$) and X-ray beam is taken into account. Fig. 2.14 and 2.15 show the translation and rotation of X-ray cone to reach the optimal pose, respectively. The cone represents the X-ray beam, the vertex of cone is the X-ray source. When the hand is out of the cone, the translation mode is firstly triggered, and the X-ray source moves along the normal direction of ROI. Otherwise, if the hand is always involved in the X-ray beam




**(a)** Result of Reconstruction.

**(b)** Labels on spinal phantom.

**Figure 2.12:** Appearance model.

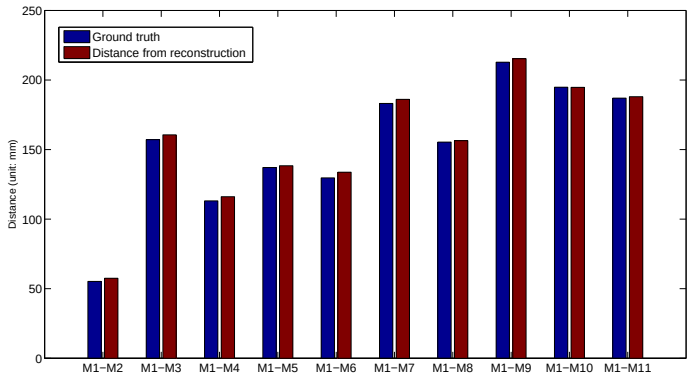


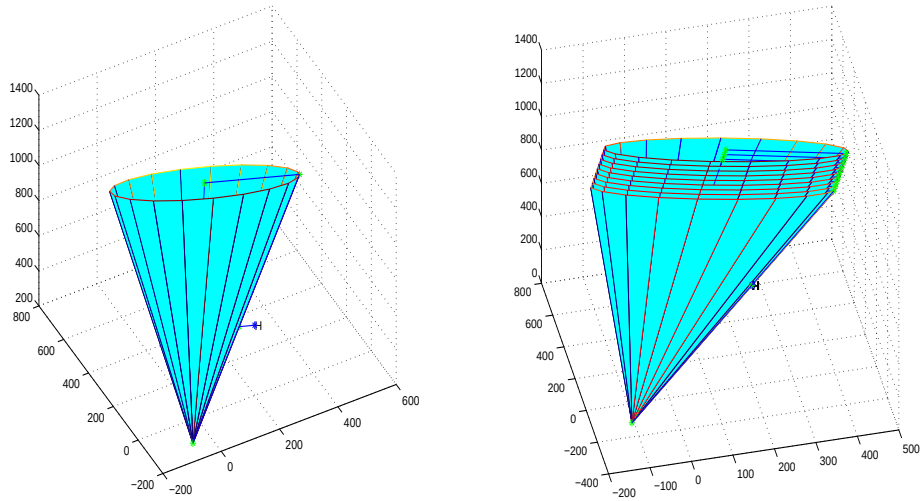

**Figure 2.13:** Accuracy of reconstruction.

in the translation mode, the cone will be rotated around the center of ROI in the plane determined by hand and axis of X-ray beam. The maximal rotational shift is set to 10° in the simulation. The rotation mode is usually skipped in practice when the reference pose is found in the translation mode.

### 2.5.3 Precision of Mobile C-arm positioning

Before the actual X-ray image capture, a floating image was generated to make sure the image is desired. To evaluate the precision of mobile C-arm positioning, the similarity between floating image and X-ray image was compared by employing correlation coefficient (CC) as the similarity metric. In this experiment, given a reference pose, the mobile C-arm was positioned using three different settings. (1) manual positioning without inverse kinematic control; (2) inverse kinematic control with 5 DoFs (base is fixed); (3) inverse kinematic control with 8 DoFs.

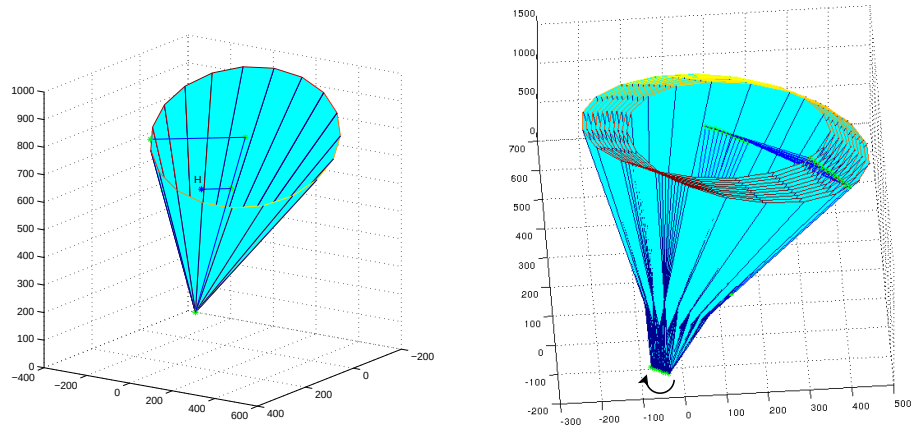
Fig. 2.16 shows the result of similarity with three setting. When the similarity is equal to 1, the floating image is identical to the corresponding X-ray image. All X-ray images are captured using one-time imaging. Using setting 1, the similarity is lower than the last two settings. Thus, the inverse kinematic control improves the precision of mobile C-arm positioning. Average similarity using setting 2 and setting 3 is similar. In setting 2, the base of mobile C-arm was fixed next to the spinal




**(a)** Initial state.

**(b)** Approximate procedure of reference pose through translation.

**Figure 2.14:** Translation mode for determining the optimal C-arm pose.




**(a)** Initial state.

**(b)** Approximate procedure of reference pose through rotation.

**Figure 2.15:** Rotation mode for determining the optimal C-arm pose.

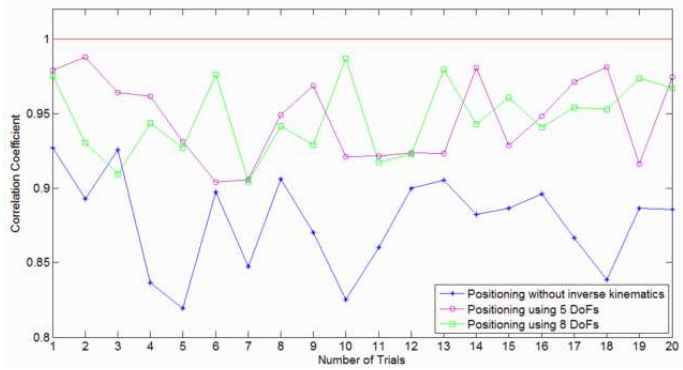



**Figure 2.16:** Precision of mobile C-arm positioning.

phantom. 5-DoF modeling is enough to position the mobile C-arm in the reference pose. In setting 3, inverse kinematic control with 8 DoFs allows the positioning of mobile C-arm from anywhere in the tracking volume of Vicon, since the movements of mobile C-arm were tracked by Vicon. 8-DoF modeling enlarges the working space of C-arm with the high precision.

## 2.6 Related Work

Some solutions and systems with different configurations are developed to position mobile C-arms precisely for capturing the fluoroscopic image with less additional radiation doses to the patient and operation time. But there is less work reporting methods in regard to the determination of C-arm pose with respect to the patient.

In Navab et al. (2006); Klein et al. (2007), Camera-Augmented Mobile C-arm (CAMC) is configured to track the source pose by the on-board camera based on the visible markers amounted on the patient's skin. And an interactive 3D model of C-arm is used in the guidance system. However, only C-arm model cannot provide an unambiguous operations for the surgical staff in the OR without considering the congested ORs with several surgical objects, such as clinical table, moving carts for instruments. Therefore, a complete interface need to be provided to interact with surgeons. Wang *et al.*Wang et al. (2012b) presents a modeling method which

integrate both the mobile C-arm and patient's table as a kinematic chain without constraining table position. That in turn means the base of C-arm can be translated along the table as the sixth DoF. However, considering the joint restrictions and the self rotation of C-arm, more degrees of freedom are required to enable the larger working space in the operation room. Without the use of markers or other prior assumptions, Schaller *et al.* Schaller et al. (2009) use a Time-of-Flight (ToF) sensor to achieve inverse C-arm positioning for interventional procedures, and mainly focus on the identification of anatomical regions on the patient body to figure out the pose of C-arm. But less details on the repositioning procedure are talked about. In Binder et al. (2006), a robotized C-arm is equipped with encoders and motors for each joint, which enables the automatic inverse kinematics. And a closed-form solution for the inverse kinematics have been found for automated C-arm positioning Matthaus et al. (2007). However, it only provides a unique solution for the C-arm joints according to the position and orientation of the X-ray beam, not accounting for the obstacle collision situations. For configurations above, all joints have mechanical limitations that dramatically restrict the working space at a desired pose to acquire fluoroscopic images. Therefore, additional DoF are necessary to increase the working space and the possibility of avoiding obstacles.

When the positioning of mobile C-arms performs, a tracking approach has to be implemented to guide the increment of each joint, and check if the X-ray source reaches the target. Traditionally, C-arm pose tracking is divided into two categories. One method is external tracking using one or more optical cameras to recover the pose. Navab *et al.* Navab et al. (2006) calculated the C-arm's pose and its required displacement for positioning using the CAMC's optical camera based on additional visible markers on patient's skin. However, C-arm becomes cumbersome with a camera and mirrors attached, while itself is unstable, when the C bracket is rotating especially. In Ladikos et al. (2008), Ladikos *et al.* proposed a real-time 3D reconstruction system using 16 optical cameras mounted on the ceiling of the interventional room for interventional environments. It can track the objects and

predict collisions by building a 3D representation of all the objects in the room, increase the operating safety and allow faster device operation. The total system cost is high, and cost-effective devices have to be found.

The other category is achieved by sensors, such as accelerometer and laser. Grezda *et al.* Grzeda and Fichtinger (2010) used tilt sensing accelerometers for rotational encoding of C-arm to track the C-arm primary and secondary angle (angular and orbital) rotations during the surgery. For comparison, a webcam is used to obtain ground-truth C-arm poses. A C-arm Laser positioning device is developed to facilitate percutaneous renal access Ko and Razvi (2007). The sensor is attaches to the side of the image intensifier of the C-arm. The aiming beam projects a parallax-free crosshair onto the patient's skin along the beam direction of mobile C-arm. That allows the radiology technician to correctly position the C-arm with the minimally unnecessary radiation exposure. However, such methods only track the position of X-ray tube/image intensifier, while the joint status and base movement are also important during the C-arm positioning. For C-arm positioning, we need to know the pose of C-arm. But previous methods are only concerned about the target position. Therefore, the practical tracking system should track the pose of whole body of the mobile C-arm.

## 2.7 Summary

A mobile C-arm is widely used for minimally invasive surgery, where C-arm positioning is a routine surgical task to take X-ray images at a desired projection view. To reduce the unnecessary radiation dose to the patient and surgeons, the precision of C-arm control is critically important for the interventional imaging techniques. Therefore, a novel configuration is proposed for mobile C-arm positioning with the aid of Virtual Reality, and additional degrees of freedom definitely enable a larger working space than the standard model. During the positioning, the surgeon can check the joint movement naturally from the outside view without self-explaining requirement.

Surgeons' hands are tracked to avoid the radiational zone. Experimental results showed that additional DoFs enable a larger working space than the standard model. The proposed system definitely facilitates and speeds up interventional procedures. In future, the proposed system will be performed on the animals. The complex structure of bones, especially the deformable spine, has to be considered. The current positioning system will be optimized in a cluttered environment to reach the higher requirement.

# Chapter 3

# Robust and Fast Initialization for Intensity-based 2D/3D Registration

# Abstract

Intensity-based 2D/3D registration is a key technique using digitally reconstructed radiographs (DRRs) to register the pre-operative volume to the patient setup during the operation. Although DRR-based method provides a high accuracy, the small capture range hinders its clinical use. In this paper, such problem was addressed by a robust and fast initialization method using a two-level scheme including automatic tracking based initialization (Level I) and multi-resolution estimation based on central-slice theorem and phase correlation (Level II). It provided almost optimal transformation parameters for intensity-based registration. Experiments using a public gold standard data set and a spinal phantom have been conducted. The mean target registration error (mTRE) was limited in the range from 2.12 mm to 22.57 mm after tracking based initialization. The capture range based on level II only was 20.1 mm, and the mTRE in this capture range was $2.92 \pm 2.21$ mm. The intensity-based 2D/3D registration using proposed two-level initialization achieved the successful rate of 84.8% with the average error of 2.36 mm. The experimental results showed that the proposed method yielded the robust and fast initialization for intensity-based registration methods. In a similar way, it can be applied to other registration methods to enable a larger capture range and robust implementation.

## 3.1 Introduction

Image-guided surgery (IGS) is a medical procedure, where 3D pre-operative volume (e.g., CT/MRI) and 2D intra-operative X-ray images usually generated by a mobile C-arm are involved. Pre-operative volume data (moving image) is used for diagnosis and surgical planning, while intra-operative images (fixed images) provide surgeon with updated information relative to the current pose of patient. By aligning these data, the planned trajectory of tools and ultimate target will be visualized with the latest information about region of anatomy, so that surgeons can confirm the surgical outcome in timeCleary and Peters (2010). Additionally, pre-operative data provides the 3D information that 2D X-ray images lack. Therefore, 2D/3D registration is performed as a key technique in IGS to estimate the transformation between the coordinates of the preoperatively acquired data and the patient setup during the operation. Various algorithms were presented for 2D/3D registration Meyer and Wolf (1999); Balter and Kessler (2007); Jerbi et al. (2013); Hurvitz and Joskowicz (2008); Lang et al. (2012); Rohlfing et al. (2005); Reda et al. (2012); Munbodh et al. (2007); Xiaogang et al. (2013); Su et al. (2013); Gerber et al. (2013), where intensity-based 2D/3D method is extensively investigated using information contained in voxels and pixels of 3D and 2D images, respectively. However, these methods suffer the misalignment caused by the initial transformation before the optimization of 2D/3D registration. Therefore, a good initialization is necessarily required to guarantee the successful rate and decrease the number of iteration (running time) of registration.

Currently, the field of intensity-based method is dominated by digitally reconstructed radiograph (DRR) based method, where multiple fixed images are captured and corresponding simulated X-ray projection images (DRRs)Markelj et al. (2010) are generated in different directions. Although DRR-based method was implemented with the high accuracy, a major issue is the small capture range which is defined as the 95% successful rate in the order of 2 mm mean target registration error (mTRE) van de Kraats G.P. Penney D. Tomazevic Th. van Walsum and Niessen (2005).

Initial guess lying outside the capture range leads to the failure of registration. It is essentially a problem of initialization for 2D/3D registration. In this case, it is obviously impractical to make sure that initial parameters always lie in the capture range. Alternatively, another solution is to enlarge the capture range to increase the probability of successful initialization.

In this paper, a two-level initialization method is proposed including tracking based initialization (Level I) and finer estimation based on multi-resolution strategy in frequency domain (Level II). Level I yields the relative pose between 2D X-ray images, and provides good initial parameters in an automatic way for Level II. There is no requirement of manual initialization for frequency estimation. In the second level, multi-resolution strategy is used in frequency estimation based on wavelet decomposition. It allows the fast implementation and improves the robustness of estimation by starting registration from the coarse level.

## 3.2 Methods

The typical intensity-based 2D/3D registration is implemented with a small capture range and low successful rate. In this paper, given two X-ray images of an anatomical structure and the corresponding CT data, the objective was to provide a robust and fast initialization for the practical use of intensity-based registration in an automatic way.

A two-level initialization scheme was proposed in this paper, including tracking based initialization (Level I) and multi-resolution frequency estimation (Level II). Level I yielded the relative pose of two X-ray images, and provided a coarse transformation parameters for Level II. In the second level, a finer estimation was obtained in the Fourier domain based on central-slice theorem and phase correlation. Especially, wavelet pyramids were used for the rotation estimation to enable the fast implementation and robust estimation.

### 3.2.1 Level I: Tracking Based Initialization

In this section, the relative pose of X-ray images and a coarse initialization were estimated based on Vicon motion tracking system. The configuration of tracking based initialization was presented in Fig. 3.1, referred to as V-Tracker. **V**icon motion tracking system was mounted on the ceiling of room. To enable that the pose of mobile C-arm can be tracked by the motion tracking system, several infrared reflective markers were attached on the X-ray source of C-arm. A spinal phantom was used in the experiments, and the markers were also attached to calculate the coarse initial transformation. Based on the markers attached on the C-arm, it was easy to estimate the relative pose of X-ray images by tracking the focal points where X-ray images were captured.

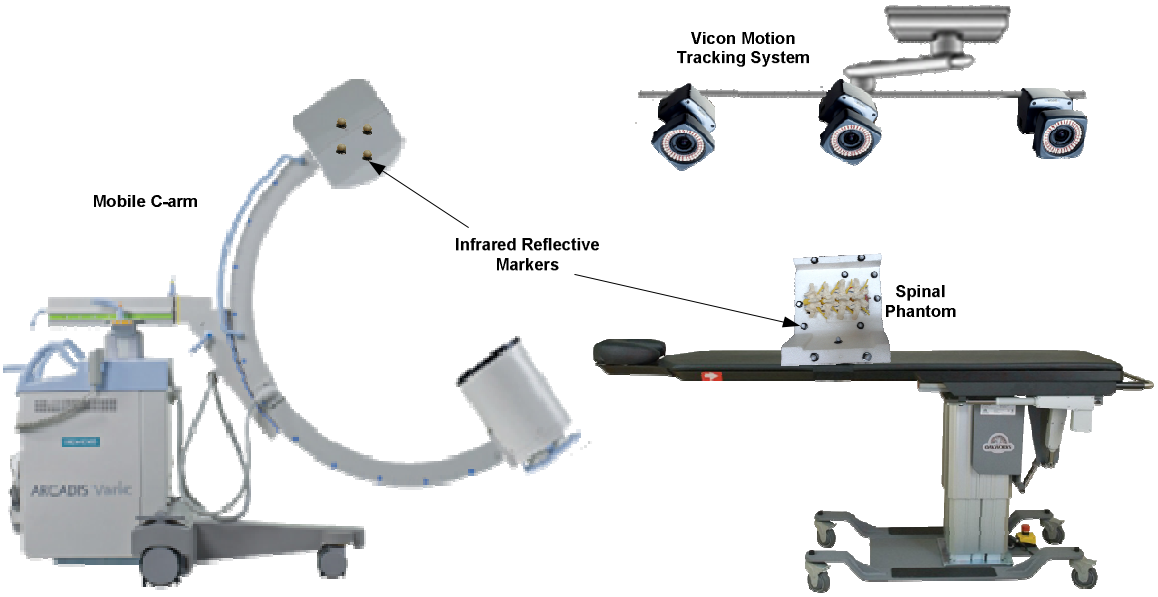



**Figure 3.1:** Configuration of tracking based initialization.

To estimate the coarse transformation of volume $T$ in the CT coordinate system, the problem is formulated as

$$T \cdot P_{i,CT} =^{CT} T_{Vicon} \cdot P_{i,Vicon} \quad i \in [1, 11] \tag{3.1}$$

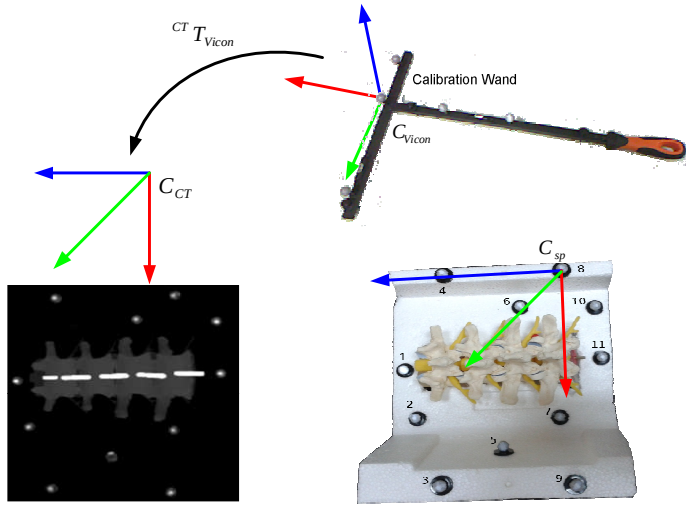



**Figure 3.2:** Illustration of the tracking-based initialization.

where $^{CT}T_{Vicon}$ is the transformation from Vicon to CT. Thanks to the clear imaging of markers in CT data, the position of marker $i$ denoted by $P_{i,CT}$ can be obtained from the CT reconstruction using the volume rendering technique in Visualization ToolKit (VTK) Ahrens et al. (2000), and $P_{i,Vicon}$ is measured by Vicon motion tracking system directly. To find the $^{CT}T_{Vicon}$, firstly we set $T$ to a identity transformation. Since the positions of markers in CT are fixed, the problem is therefore reduced to the initial placement of the spinal phantom.

As shown in Fig. 3.2, $C_{CT}$, $C_{Vicon}$ and $C_{sp}$ are the coordinate system of CT volume, Vicon motion tracking system, respectively. An off-the-shelf calibration method was employed to set $C_{Vicon}$ using a 240 mm wand (5-Marker wand). The x-y plane is determined by these markers, and the normal vector of this place is defined as z-axis. A local coordinate system $C_{sp}$ on spinal phantom is defined as the same as the CT frame. The origin is located at marker 8. The vector from marker 8 to marker 4 defines z-axis, the y-axis is defined by the normal vector of the place in the middle where markers 1,2,5,6,7,10,11 were placed. The cross product of y-axis and z-axis results in x-axis.

In our experiment, the transformation $T$ is equal to identity matrix, when three conditions were satisfied.

- the z-axis of $C_{sp}$ is parallel with the y-axis of $C_{Vicon}$,
- x-axes of $C_{sp}$ and $C_{Vicon}$ are in the opposite direction,
- When the mobile C-arm is in the vertical position (detector-over-table Reaungamornrat et al. (2012)), the line connecting the X-ray source and center of image intensifier is perpendicular to the middle plane of spinal phantom and crosses through the center of plane.

Then ${}^{CT}T_{Vicon}$ can be found by solving Eq. 3.1. Once ${}^{CT}T_{Vicon}$ is determined, The new position of every marker can be transformed into the $C_{CT}$. The transformation $T$ can be obtained based on the paired-point registration (PPR).

$$\begin{aligned} P'_{i,CT} &= {}^{CT}T_{Vicon} \cdot P_{i,Vicon} \\ T &= PPR(P'_{CT}, P_{CT}) \end{aligned} \tag{3.2}$$

### 3.2.2 Level II: Multi-resolution Rotation Estimation and Translation Estimation in Fourier Domain

Following the tracking based initialization, a frequency method based on the multi-resolution strategy is proposed to obtain more accurate and robust estimation in a faster way. As shown in Fig. 3.3, Fourier transform is applied to the pre-operative volume and X-ray images, followed by the wavelet based pyramids, which allows a rough estimation using low-resolution image in the following optimization, and subsequently the parameters are refined when the resolution is increased gradually. The estimation at high resolution is initialized by a reasonable estimate from the coarse resolution. Rotation and translation estimation can be separated in Fourier domain. It will be explained in detail later. Firstly, the rotation estimation (procedures in the red dashed rectangle) is implemented iteratively based on the

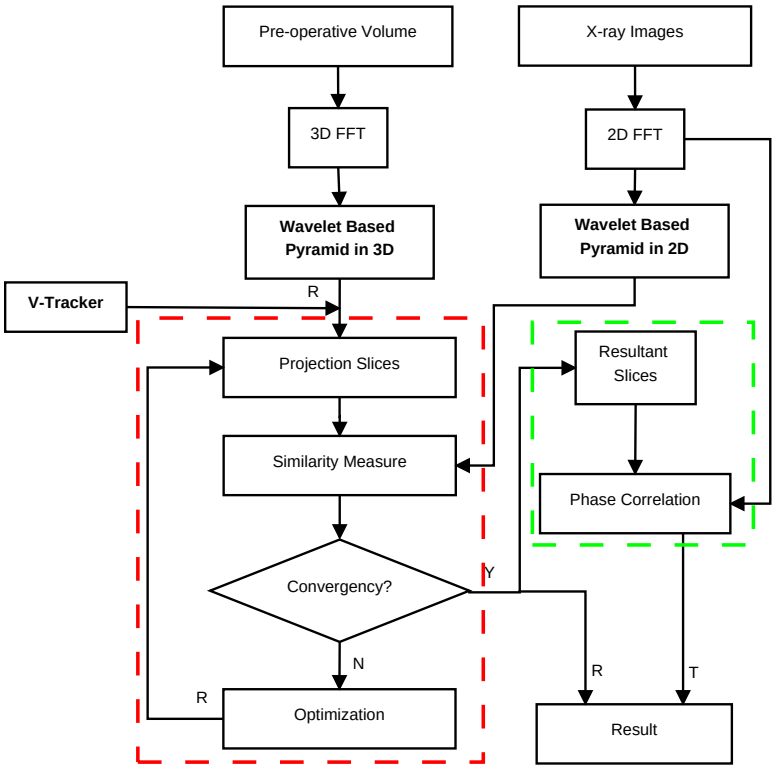



**Figure 3.3:** Multi-resolution estimation in Fourier domain.

initial parameters from V-Tracker by starting from the coarse level of wavelet based pyramid. In each iteration, the dimension of search space is reduced to 3, while it is 6 in the traditional intensity-based 2D/3D registration. Once the optimal rotation is obtained, translation vector is estimated in a non-iterative way using phase correlation method.

To explain the separation of rotation and translation estimation, let $\Phi_2$ be a representation of $\Phi_1$ in 3D after a rigid transformation $(R|T)$ is applied. The relationship between $\Phi_1$ and $\Phi_2$ can be written as

$$\Phi_2(x, y, z) = \Phi_1(R(x, y, z) + T). \tag{3.3}$$

Applying Fourier shift theorem to Eq. 3.3,

$$\hat{\Phi}_2(u,v,w) = \hat{\Phi}_1(R(u,v,w)) \cdot \exp(2i\pi(t_x u + t_y v + t_z w)), \tag{3.4}$$

where $\hat{\Phi}$ is the result after the Fourier transform. $(u, v, w)$ is a point in Fourier domain, and $(t_x, t_y, t_z)$ are the elements in $T$. The modulus is expressed in term of rotation, and the phase part is related to translation information. Therefore, we can estimate the rotation and translation separately. In this paper, the versor transformation in Insight Segmentation and Registration Toolkit (ITK) Yoo et al. (2002) is used to model the rotation. To simplify the translation estimation, X-ray images were captured in the Anterior-Posterior (AP) view and Lateral (LAT) view using the mobile C-arm in the experiments, and the desired views were achieved based on the Vicon motion tracking system.

### Wavelet pyramid in Fourier domain

Multi-resolution strategy in Fourier domain was the key concept in Level II, and used to optimize the rotation parameters. In this paper, pyramids for 2D and 3D data were constructed based on wavelet transform (WT). Compared with Gaussian and Laplacian pyramid, wavelet pyramid is achieved in one step for less computing cost by merging filtering and downsampling steps together Xu and Chen (2007). It is more efficient in Fourier domain especially Adelson et al. (1990). According to the convolution theorem, wavelet transform is the multiplication of signals and the mother wavelet with a set of scales and translational values in frequency domain. Unlike the convolution operation, it allows the simultaneous computation.

In this paper, Quadrature Mirror Filter (QMF) Akansu et al. (1993); Mueller et al. (2006) as a discrete orthogonal form of wavelet was used. Three levels of resolution were used, while the level for WT was set to 2. Fig. 3.4 shows the pyramid of fixed

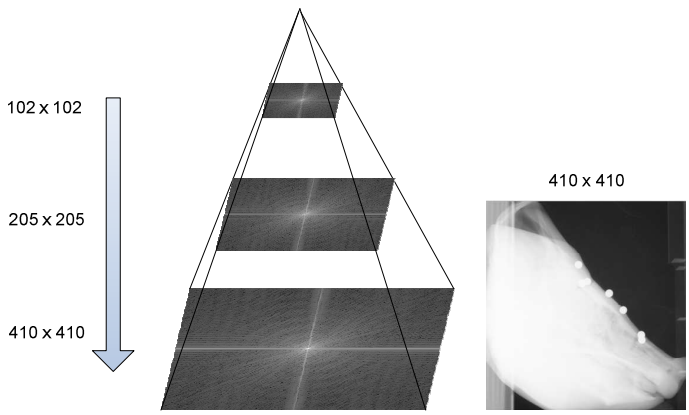



**Figure 3.4:** Wavelet pyramid of the image in golden standard data set Pawiro et al. (2011).

image in Fourier domain. The right column is the origin image with the resolution of $410 \times 410$.

### Rotation estimation

Central-slice theorem was applied to generate the projection images which were used to optimize the rotation parameters. It defines a rotational relationship between

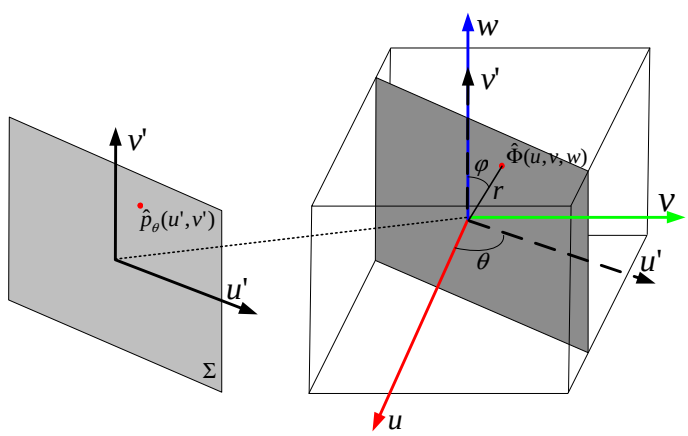



**Figure 3.5:** Illustration of central-slice theorem in 3D.

the Fourier spectrum of an N-dimensional function and Fourier spectra of its (N-1)-dimensional projections. In this paper, we will get 2D projections, when the pre-operative data is in 3D. As shown in Fig. 3.5, the parallel projection $p_\theta(u', v')$ of an object $\Phi(x, y, z)$ on a plane $\Sigma$ through the origin with angle $\theta$ relative to the x-axis evaluated at a distance $r$ from the origin and angle $\varphi$ relative to z-axis can be written as

$$\begin{aligned} p_\theta(rsin\varphi, rcos\varphi) = \iiint \Phi(x, y, z)\delta(xcos\theta sin\varphi + ysin\theta sin\varphi \\ + zcos\varphi - r)dxdydz, \end{aligned} \tag{3.5}$$

where $\delta$ is the Dirac delta function. According to central-slice theorem,

$$\begin{aligned} \hat{p}_\theta(u', v') &= \hat{\Phi}(u, v, w) \\ &= \hat{\Phi}(rcos\theta sin\varphi, rsin\theta sin\varphi, rcos\varphi) \\ u' &= rsin\varphi, v' = rcos\varphi, \end{aligned} \tag{3.6}$$

where $\hat{p}$ is the 2D Fourier transform of projection image, and $\hat{\Phi}$ is 3D Fourier transform of the volume. Thus, we can obtain projection image through the volume by extracting the slice through $\hat{\Phi}$ at the same angles.

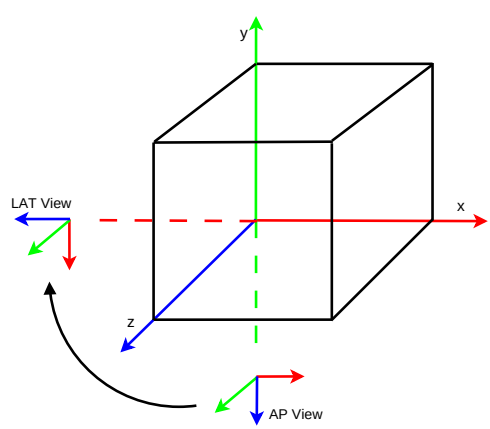



**Figure 3.6:** Geometry of AP and LAT views.

In this paper, two X-ray images were used as fixed images for 2D/3D registration. To resample the slice in a specific direction through the origin of volume in the Fourier domain during the optimization, rotation matrix has to be constructed. Inspired by the camera transformation in OpenGL, rotation matrix is calculated as

$$R = \begin{bmatrix} s_x & u_x & -f_x \\ s_y & u_y & -f_y \\ s_z & u_z & -f_z \end{bmatrix}, \tag{3.7}$$

where $f$ is the direction of projection, $u$ is the up vector, and $s$ is the cross product of $f$ and $u$. As shown in Fig. 3.6, the rotation matrices in AP and LAT view can be constructed. In the AP view, the projection plane is parallel with x-z plane in the volume coordinate, while the image in LAT view is generated by projection along axis $x$. The relation between both matrices is

$$R_{LAT} = R_{AP} * \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}. \tag{3.8}$$

For each iteration of rotation optimization, two 2D images in Fourier domain were extracted from a specific level of wavelet pyramid according to

$$\begin{aligned} Slice_{k,AP} &= P_{AP}(\hat{\Phi}_k) \\ Slice_{k,LAT} &= P_{LAT}(\hat{\Phi}_k), \end{aligned} \tag{3.9}$$

where $Slice_k$ and $\hat{\Phi}_k$ are the projection slice and 3D volume at level $k$ in Fourier domain.

The similarity measure between projection slices and intra-operative X-ray images $I$ can be expressed by combining both measures in AP and LAT views.

$$SM = SM(Slice_{AP}, \hat{I}_{AP}) + SM(Slice_{LAT}, \hat{I}_{LAT}). \tag{3.10}$$

Please note that Fourier transform is applied to $I$, since $Slice_{Ap}$ and $Slice_{LAT}$ are presented in Fourier domain. In this paper, normalized gradient correlation was employed to calculate the similarity measure, and optimization was implemented by Powell's method.

### Translation estimation

Once the rotation is determined, we can estimate the translation according to Eq. 3.4 using the phase correlation method (PCM), which is a registration technique based on the Fourier shift theorem. Let $Slice_{AP}$ and $Slice_{LAT}$ be the resultant projection slices of rotation estimation in x-z and y-z plane, respectively. Substitute the results into Eq. 3.4,

$$\begin{aligned} Slice_{AP}(u, w) &= \hat{I}_{AP}(u, w) \exp(2i\pi(t_x u + t_z w)) \\ Slice_{LAT}(v, w) &= \hat{I}_{LAT}(v, w) \exp(2i\pi(t_y v + t_z w)). \end{aligned} \tag{3.11}$$

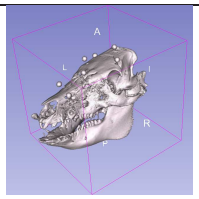
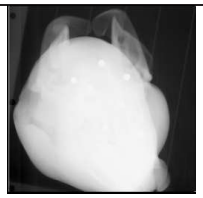
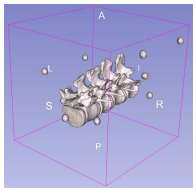
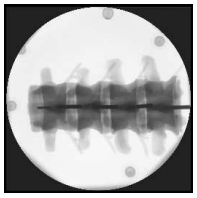
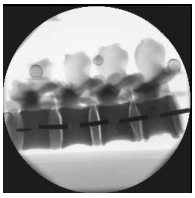
Let's take the AP view for example, the phase difference between $Slice_{AP}(u, w)$ and $\hat{I}_{AP}(u, w)$ can be calculated using the cross-power spectrum $C(u, v)$, defined as

$$C(u, v) = \frac{Slice_{AP}(u, w)\hat{I}^*_{AP}(u, w)}{\mid Slice_{AP}(u, w)\hat{I}^*_{AP}(u, w) \mid}. \tag{3.12}$$

where $\hat{I}^*_{AP}(u, w)$ is the complex conjugate of $\hat{I}_{AP}(u, w)$. The inverse Fourier transform $c(x, z)$ of $C(u, v)$ is a Dirac delta function that indicates the relative shift $(t_x, t_z)$.

$$\begin{aligned} c(x, z) &= \delta(x - t_x, z - t_z) \\ (t_x, t_z) &= \arg\max_{(x,z)} c(x, z). \end{aligned} \tag{3.13}$$

**Table 3.1:** Summary of the data sets used in the experiments.

| Data Source | Pre-operative CT | Fixed Images | |
|---|---|---|---|
| | | AP view | LAT view |
| Golden standard data set |  |  | |
| Spinal phantom |  |  |  |

A similar process can be carried out to estimate the difference $(t_y, t_z)$ in the lateral view.

## 3.3 Experiments and Discussions

In this section, three sets of experiments were performed to validate the proposed two-level automatic initialization method. Multi-resolution rotation estimation in Level II was firstly tested using DRR images as the fixed images to determine the improvement in speed and robustness using the wavelet pyramid. The second experiment was to evaluate the complete initialization combining both levels of initialization methods. In the last experiment, the proposed initialization method followed by DRR-based 2D/3D registration was tested to verify whether it can improve the successful rate of registration compared with the results in van der Bom et al. (2010).

The experiments used the golden standard data set Pawiro et al. (2011) and the CT data of spinal phantom acquired by the CT scanner (Aquilion ONE, Toshiba, Japan). Table 3.1 summarized the pre-operative CT, fixed images and ground truth in experiments. The specifications including the resolution and spacing of the data sets were listed in Table 3.2.

**Table 3.2:** Specification of the data sets used in experiments.

| Data Source | Pre-operative CT Resolution (pixel) | Pre-operative CT Spacing (mm) | Fixed Image Resolution (pixel) | Fixed Image Spacing (mm) |
|---|---|---|---|---|
| Golden standard | 326×326 ×330 | 1.0×1.0 ×1.0 | 410×410 | 2.498×2.498 |
| Spinal phantom | 512×512 ×227 | 0.559×0.559 ×1.0 | 1024×1024 | 1.0×1.0 |

### 3.3.1 Performance Measurements

To estimate the performance of the proposed method, we used mean Target Registration Error (mTRE), capture range, running time and successful rate as the performance measurements. The mTRE was defined as the Euclidean distance between fiducial markers in the pre-operative volume transformed with the golden standard transformation (ground truth) and estimated transformation,

$$mTRE = \frac{1}{N}\sum_{n=1}^{N} \parallel Tp_n - T_g p_n \parallel, \tag{3.14}$$

where $N$ is the total number of fiducial markers $p_n$, $T$ is the result from the initial guess or registration algorithm. and $T_g$ is the gold standard transformation. In the pre-operative CT summarized in Table 3.1, there are seven and eleven markers respectively, which can be used to calculate the $mTRE$. Particularly, with the target points in the golden standard data set, a rotation of 1.0° around one of the three axes will lead to an mTRE of $1.2 - 1.6\,mm$. Following Bom's work, the registration error before the multi-resolution frequency estimation and the error after the estimation process are refer to as $mTRE_{start}$ and $mTRE_{estimate}$, respectively. $mTRE_{final}$ is the final result after intensity-based registration.

The estimation with the $mTRE$ smaller than 7.5 mm was considered successful, and the capture range was defined where the successful rate can reach 95% van de Kraats G.P. Penney D. Tomazevic Th. van Walsum and Niessen (2005). The running

time was mainly used to quantify the efficiency of initialization method compared with Bom's method, which required 7.5 min per initialization.

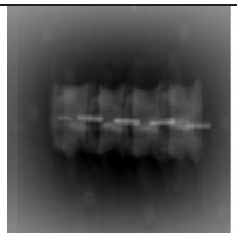
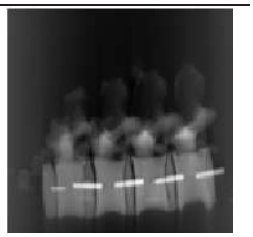
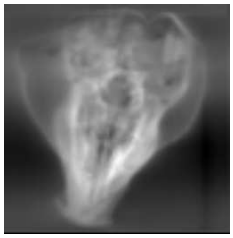
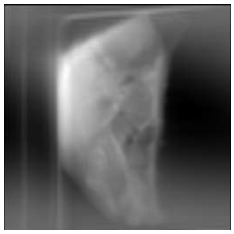
In all experiments, Powell optimizer was used for the initial estimation with the same parameters for comparison of different methods. The tolerance of similarity measure was set to 0.001, and the maximum of iterations was 100. For multi-resolution strategy, the tolerance of search step and step length were set to 0.08 and 8.0, respectively. When the resolution increased, both step parameters were decreased by 2. All experiments were implemented within ITK.

### 3.3.2 Test of Multi-resolution Rotation Estimation

In this experiment, the method of multi-resolution rotation estimation based on central-slice theorem was evaluated using the spinal phantom and golden standard data set, respectively. The projection images obtained with central-slice theorem are shown in Table 3.3. The spinal phantom was mainly used for rotation estimation using the projection images obtained with central-slice theorem and comparison of results with and without multi-resolution strategy. The performance of rotation and translation estimation was evaluated using the golden standard data set. For both data sets, DRRs were generated with known various rotation and translation parameters as the ground truth to register to the corresponding CT data. With these experiments, mTRE, capture range and running time of multi-resolution rotation estimation were quantified.

In the experiment using the spinal phantom, the ground truth of rotational parameters is $R = (R_x, R_y, R_z) = (90°, -90°, 0°)$, while the translation parameters were fixed at 0 mm. The initial parameters for rotation estimation varies in the range of $[92.82°, 104.28°]$, $[-89.95°, -78.5°]$ and $[0°, 11.46°]$ for x-, y- and z-axis, respectively. The same set of parameters were used to evaluate the performance without and with the multi-resolution strategy. There were 250 trials performed in total in this experiment.

**Table 3.3:** Projection images obtained with central-slice theorem.

| Data Source | AP View | LAT View |
|---|---|---|
| Spinal phantom |  |  |
| Golden standard |  |  |

The results are shown in Fig. 3.7. The red line in Fig. 3.7a and 3.7c indicates the threshold for the successful estimation (7.5 mm), and the boundary of improvement is indicated by the blue line. Compared with the result using the method in Bom's work, multi-resolution rotation estimation was more robust, especially when $mTRE$ was less than 17.5 mm. Also, the running time was faster with the average time of 464.8 s for each estimation, while the 529.9 s was taken without using multi-resolution strategy.

Using the golden standard data set, multi-resolution frequency estimation was performed, including rotation and translation estimation. Fig. 3.8 shows the results with the rotation varying from $-15°$ to $15°$ and translation offsets in the range from -10 to 10 mm for all three axes. The results above the blue line indicate no improvement, and ones under the red line (7.5 mm) denote the successful estimation. The capture range of the frequency estimation was 20.1 mm. The mTRE in this capture range was $3.3 \pm 1.2$ mm, while the standard deviation of mTRE in Bom's work was 1.6 mm. Thus, the frequency estimation using wavelet pyramid is more robust.

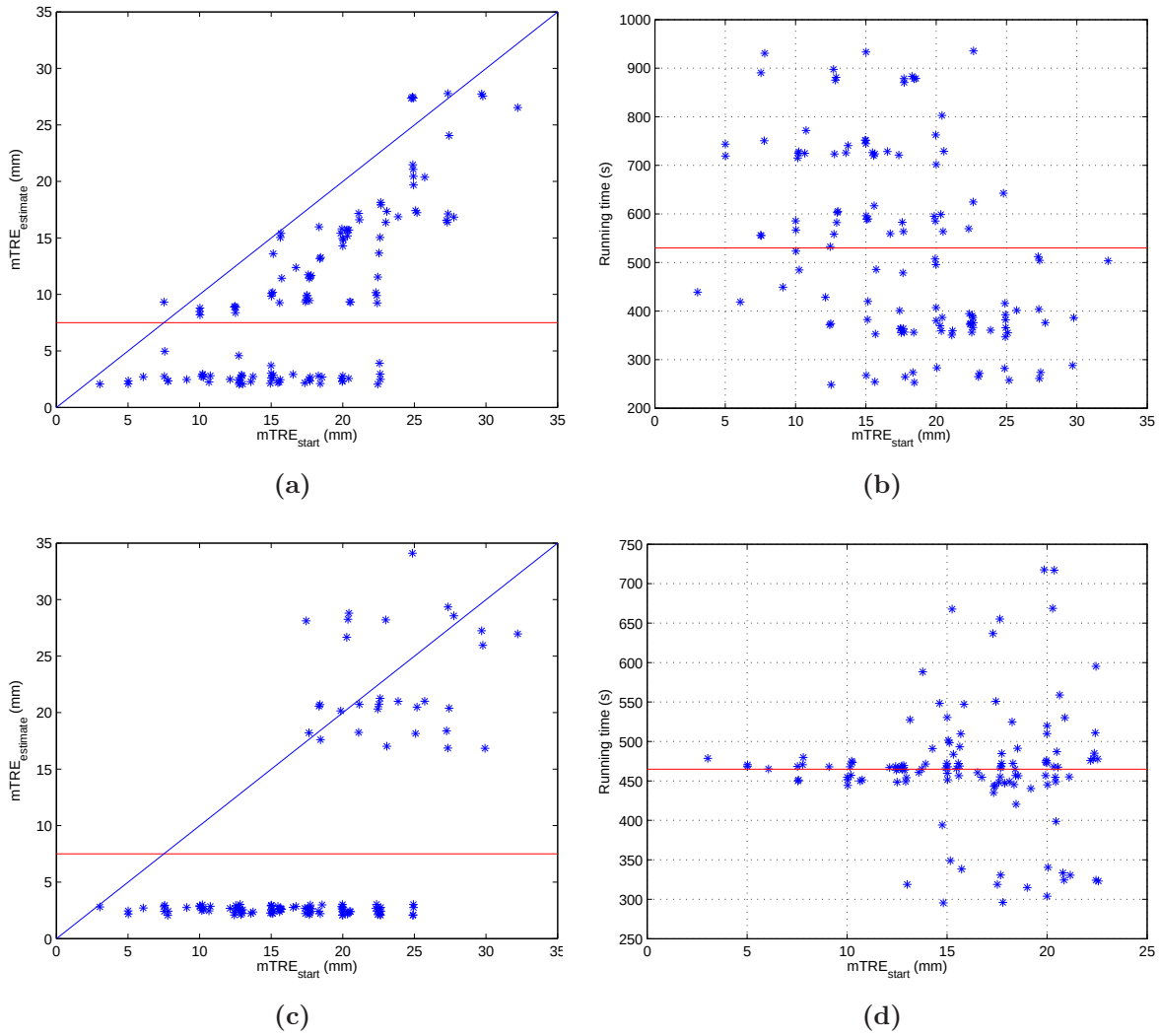



**Figure 3.7:** Rotation estimation based on multi-resolution strategy using spinal phantom. (a) and (b) represent the mTRE and running time without using multi-resolution strategy, respectively. (c) and (d) are the corresponding performances using multi-resolution strategy.

### 3.3.3 Test of Complete Initialization

In this experiment, the combination of tracking based initialization and multi-resolution frequency estimation was evaluated using the spinal phantom. The pose of phantom was tracked by Vicon motion tracking system and changed with the rotation in the range of $[-20°, 20°]$ and translation offsets from $-10$ to 10 mm along

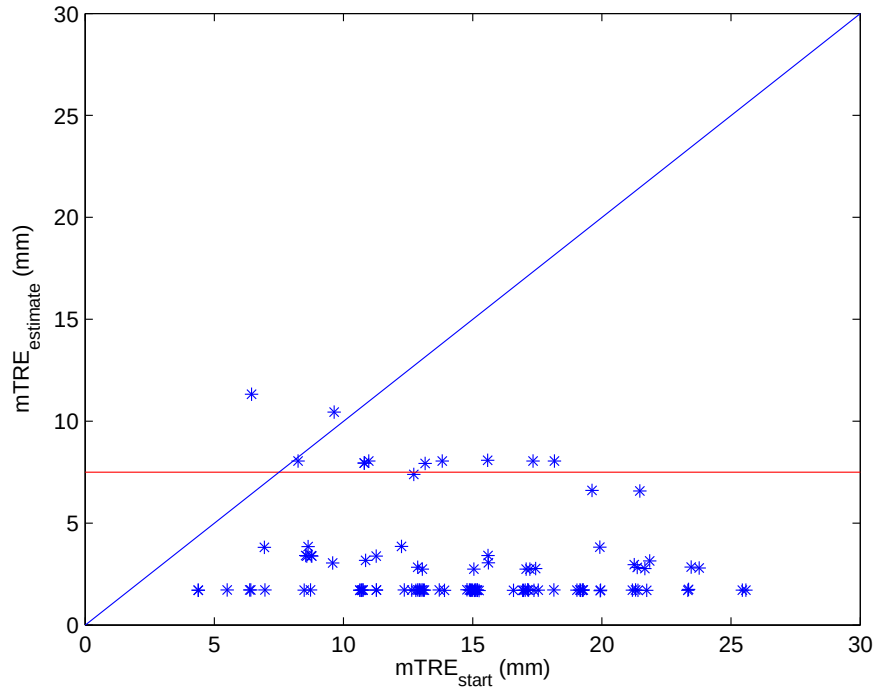



**Figure 3.8:** Frequency estimation using golden standard data set.

three axes, respectively. Two X-ray images were acquired from the AP view and later view, respectively. The ground truth of transformation parameters was determined by 2D/3D intensity-based registration with elastix toolbox Klein et al. (2009). The mask images were used to match the circle shape of the actual X-ray images and reduce the influence of sharp edges.

Fig. 3.9 shows the results using the complete initialization. $mTRE_{start}$ was located in the range from 2.12 to 22.57 mm provided by the tracking based initialization. This uncertainty was mainly caused by the inaccurate measurements of markers' positions in the CT data and placement of spinal phantom when estimating the transformation ${}^{CT}T_{Vicon}$. The method of complete initialization provided the improvement for all experiments. In 85.95% of cases, the mTRE after estimation was less than 7.5 mm. And the mean and standard deviation of $mTRE_{estimate}$ for the successful estimation were 2.38 mm and 2.16 mm, respectively.

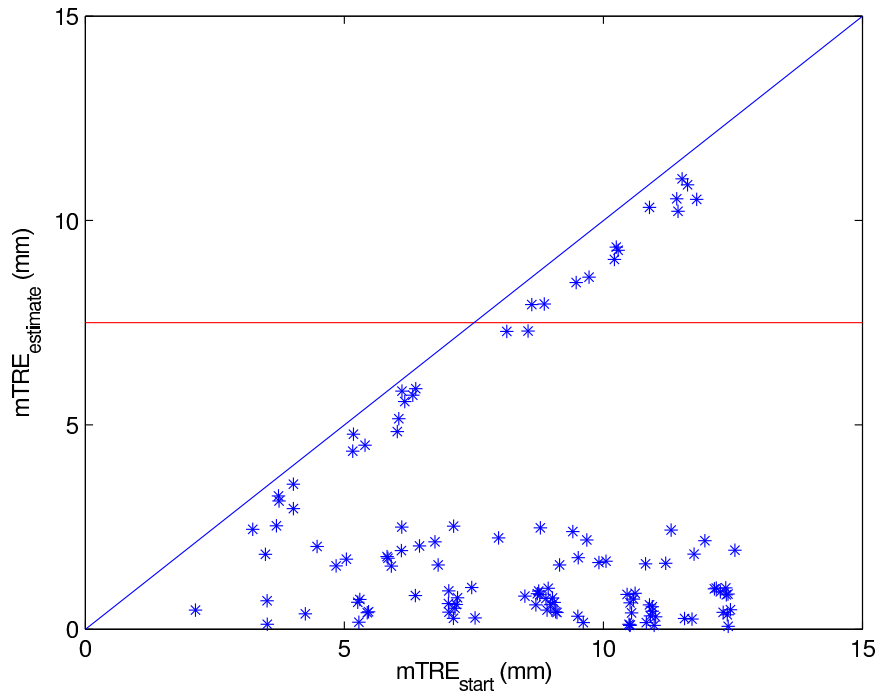



**Figure 3.9:** Results of the complete initialization test.

### 3.3.4 Test of Intensity-based Registration Using Complete Initialization

2D/3D intensity-based registration using the complete initialization was tested using the spinal phantom. There were three sets of experiments performed in this section. The traditional registration (IBR) was performed using the normalized gradient correlation as the similarity measure, and optimization was implemented by Powell's method. The ranges of rotation and translation in Section 3.3.3 were applied to provide the initial parameters in this experiment. The registration following frequency estimation (F-IBR) and complete initialization including Level I and Level II (TF-IBR) were also implemented, respectively.

Fig. 3.10 shows the $mTRE_{final}$ of IBR (blue triangles), F-IBR (green triangles) and TR-IBR (red circles). The red line ($mTRE_{final}$=2 mm) indicates the successful registration. Only 1 to 2 iterations were required using TR-IBR, while the registration using IBR took 30 iterations on average with the maximum of 75. And the successful

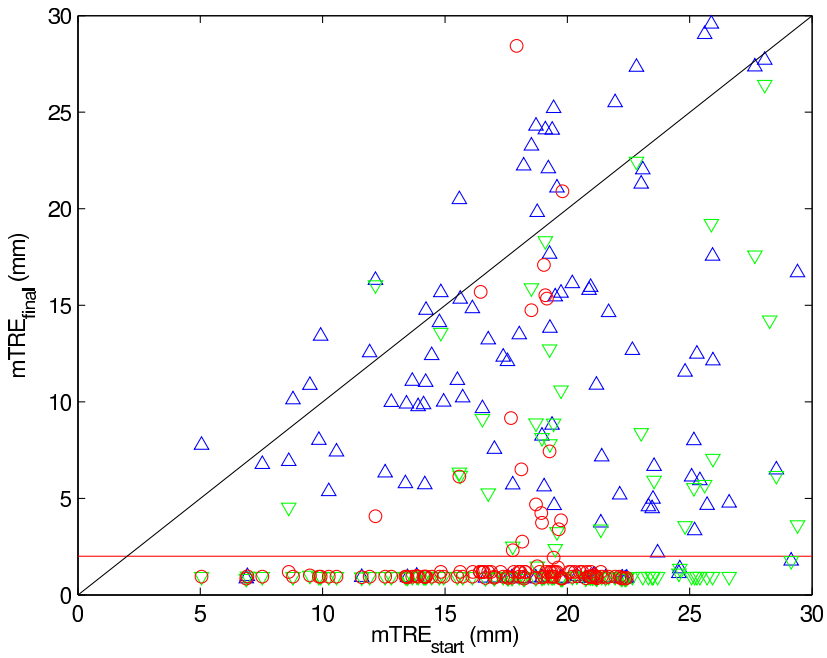



**Figure 3.10:** Results of 2D/3D intensity-based registration.

rate of TF-IBR was improved greatly from 72.8% to 84.8%, when tracking based initialization was used. The average $mTRE_{final}$ was 2.36 mm, while 3.1 mm was reported on $mTRE_{final}$ in Bom's work.

## 3.4 Related Work

A robust and fast initialization for intensity-based 2D/3D registration was performed in two steps, including tracking based initialization and multi-resolution rotation estimation. In this section, C-arm pose tracking and pyramid methods are discussed as two key techniques which are related to the contributions of this paper.

### 3.4.1 C-arm Pose Tracking

In the image-guided surgery, the mobile C-arm is commonly used as its compact structure and flexible functions. For 2D/3D registration, to estimate the relative

pose between intra-operative X-ray images and position the C-arm at a desired pose, mobile C-arm has to be tracked, especially the x-ray tube.

The approaches to estimate the relative pose between X-ray images were mainly classified into two categories. One was using an X-ray opaque fiducial marker which was imaged together with the anatomy of interest. Several specially structures were designed to estimate the pose of C-arm from the X-ray images Ayad et al. (2010); Chintalapani et al. (2008); Jain and Fichtinger (2006); Jain et al. (2005); Otake et al. (2012). In Ayad et al. (2010), a fiducial composed of a set of coplanar ellipses was used to track the C-arm since a 3D ellipse projects to an ellipse in the image. Chintalapani *et al.* Chintalapani et al. (2008) used the similar method, and showed that a single ellipse and a point correspondence can achieve the estimation of C-arm pose. In Jain et al. (2005); Jain and Fichtinger (2006), a single-image-based fluoroscope tracking (FTRAC) method was presented using an external fiducial consisting of a set of ellipses, lines, and points, which were mathematically optimized. A hybrid fiducial was designed by Otake *et al.* Otake et al. (2012). It contains two parts, including one fiducial that is visible to the optical tracking, and the other one is FTRAC, which can be seen in the X-ray image. In Reaungamornrat et al. (2012), Reaungamornrat *et al.* presented an on-board surgical tracking system within which a video-based tracker was mounted on the gantry of a mobile C-arm next to the flat-panel detector. It allowed the tracking of object from various C-arm angulations. In the same way, the pose of mobile C-arm can be monitored in turn.

The other category for the C-arm tracking was using an external tracker. Electromagnetic and optical trackers are frequently used in the most of IGS applications Cleary and Peters (2010); Livyatan et al. (2003); Sadowsky et al. (2007). However, sensitivity of electromagnetic trackers to the metal objects in the operating room can distort the electromagnetic field and reduce the accuracy. In this paper, optical tracker (Vicon motion tracking system) was used. The calibrated tracking volume is large enough to cover the whole working space, even though the mobile C-arm is moving.

### 3.4.2 Pyramid Representations

In order to avoid the false local optima, improve the robustness of convergence and reduce the running time, a hierarchical multi-resolution strategy (image pyramid) is usually used. Image pyramid is typically represented based on the two steps consisting of filtering and downsampling, e.g. Gaussian pyramid. Another way to construct the pyramid is using wavelet transform, which is implemented in one step leading to less computing cost. Several methods of 2D/3D registration were reported using multi-resolution strategy for optimization. In Kubias et al. (2007), Gaussian pyramid was used for a extended global optimization strategy by combining with a multi-scale method. Xu *et al.* Xu and Chen (2007) proposed a wavelet-based multi-resolution strategy. It allowed a hybrid metric by combining the mutual information and spatial information from the low-frequency coefficient pair and three high-frequency coefficient pairs,respectively. Recently a multi-resolution registration method was presented using firefly algorithm and Powell's method Xiaogang et al. (2013). The multi-resolution strategy was used based on wavelet transformation. The firefly algorithm was performed in the low level, while in the higher level, Powell's method was adopted to obtain the finer result. Besides, Rohlfing *et al.* Rohlfing et al. (2005) proposed a progressive attenuation field (PAF) to generate DRR images, and multi-resolution DRR image was obtained by sampling PAF parameters. However, all these methods were implemented in the intensity domain based on different pyramid representations.

In Munbodh et al. (2007), a part of work was achieved in frequency domain to find a frequency band in which the information of bone anatomy was better expressed than non-bony anatomy. But the following processing was still performed in the intensity domain based on Gaussian pyramids for the optimization. In this paper, all processes were performed in Fourier domain. That is, the wavelet-based pyramids were frequency representations. Considering the convolution theorem in

Fourier domain, the wavelet operator was more efficient than the implementation in the intensity domain.

## 3.5 Discussion and Summary

Intensity-based 2D/3D registration is a critical problem to register the pre-operative volume to the patient setup in the image guided surgery. Currently the registration is dominated by DRR-based method with a high accuracy. However, the small capture range limits its use in the clinical application. In this paper, a two-level scheme was proposed to address such problem by a hardware-software integrated approach. In the first level, automatic tracking based initialization based on Vicon motion tracking system estimated the relative pose of X-ray images and generated a coarse transformation parameters for the next level. A frequency estimation using multi-resolution strategy provides almost optimal transformation parameters for intensity-based 2D/3D registration. A public golden standard data set and a spinal phantom were used in experiments. The capture range using frequency estimation only was enlarged to 20.1 mm, and the mTRE in this capture range was $2.92 \pm 2.21$ mm. Two-level estimation has the successful rate of 84.8% with the average $mTRE_{final}$ of 2.36 mm.

Compared with the results in Bom's work, the performances of initialization for intensity-based registration were improved, including the mTRE, capture range, successful rate and running time. Especially for the running time, the multi-resolution strategy speeded up the frequency estimation by 65.1 s. Although it was not improved greatly, more robust implementation was achieved. Further more, the implementation of proposed initialization method can be accelerated by the parallel computing based on GPU Ruijters et al. (2008); Otake et al. (2012).

# Chapter 4

# Geometry Constrained Sparse Embedding for Multi-dimensional Transfer Function Design in Direct Volume Rendering

# Abstract

Direct volume rendering (DVR) is commonly employed for the medical visualization. Multi-dimensional transfer functions are used in DVR to emphasize the region of interest in details. However, it is impractical to interact directly with the functions in more than three dimension. This paper proposes a novel framework called geometry constrained sparse embedding (GCSE) for dimensionality reduction (DR). GCSE allows the conventional DR methods to be applied to a dictionary with much smaller atoms instead. The mapping derived by the dictionary feeds to the original features to obtain the ones in the reduced dimension. To obtain a good dictionary, the intrinsic structure of the features is encoded in the sparse embedding based on a geometry distance. In addition, stochastic gradient descent algorithm is employed to speed up the dictionary learning. Various experiments have been conducted using both synthetic and real CT data sets. Compared with conventional methods, GCSE not only produces the similar results, but also performs well with the capability to handle the large data set more powerfully. The rendering results using the real CT data have demonstrated the effectiveness of GCSE.

## 4.1 Introduction

In the image-guided surgery (IGS), 2D/3D registration is a key technique to align the pre-operative volume in 3D (e.g. CT and MRI) and intra-operative X-ray images. Efficient interpretation and perception of the registered data is crucial to surgeons during the clinical intervention Wang et al. (2012a), especially for the 3-D pre-operative volume. Commonly the direct volume rendering (DVR) technique is employed for 3-D medical visualization by projecting the volume data in a projected image without the data misclassification in the segmentation stage Cleary and Peters (2010). DVR is also used in other procedures, such as surgical training Georgii et al. (2007), diagnosis and surgical planning Tietjen et al. (2005); Beyer et al. (2007).

In order to define the meaningful visualization and emphasize the regions of interest in the volume data, transfer functions (TFs) are chosen for mapping from voxels with the scalar information to color and opacity properties Pfister et al. (2001); Kindlmann (2002). The simple type of TF is designed based only on the scalar value per voxel, but it can not discriminate two parts with similar intensity properly. To address this problem, multi-dimensional TFs are proposed using feature vectors consisting of several features, such as gradient magnitude, directional second derivative, curvature and statistical measures de Moura Pinto and Freitas (2007); Kim et al. (2010). In this way, more accurate perception can be provided for surgeons and better outcome is enabled in further processing (e.g. data classification and object recognition).

As shown in Fig. 4.1, multiple features in TF domain are extracted, and stored in a vector every voxel. As the dimensionality of TF domain increases, the direct interaction with the transfer function becomes impractical. In this case, we have to find a low-dimensional representation of high-dimensional features to enable the effective interaction. One solution to this problem is dimensionality reduction (DR) which is the transformation from the high-dimensional data to a representation in the reduced dimension without significant information loss Van der Maaten et al. (2009).

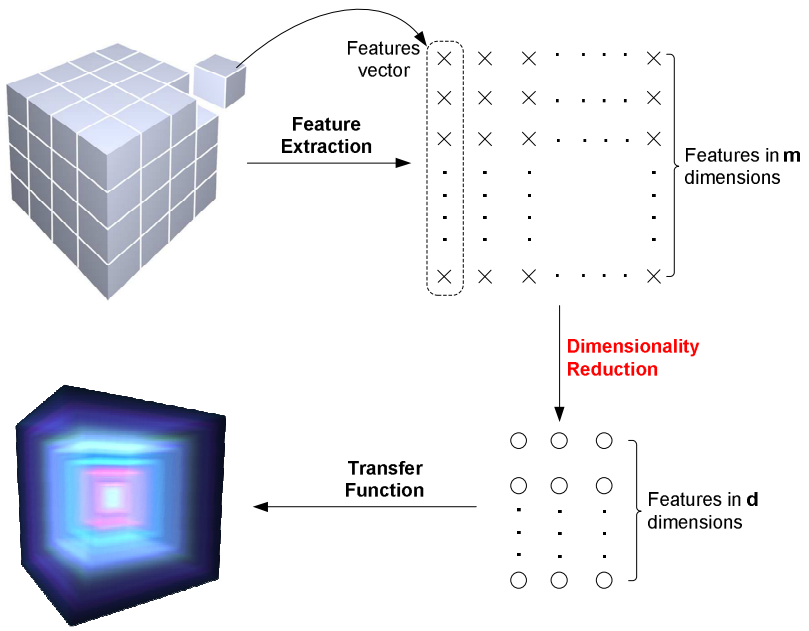



**Figure 4.1:** Dimensionality reduction for transfer function design.

Another advantage of dimensionality reduction is that the redundant information can be filtered out simultaneously. Therefore, an approach of dimensionality reduction is employed to facilitate the design of transfer function in this paper.

In this paper, a novel framework of Geometry Constrained Sparse Embedding (GCSE) is proposed. To improve the efficiency of dimensionality reduction, the mapping matrix is separated from the framework of SR, and it is obtained by applying the conventional method (i.e. PCA, Isomap and LLE) to the dictionary under the assumption that if the features vectors can be described well over a overcomplete dictionary, the mapping derived from the dictionary can embed the entire features in the target dimension. Therefore, the problem is reduced to the optimization of dictionary. To this end, the intrinsically geometrical structure based on a geometry distance are taken into account. The geometry distance is to measure the similarity and find the geometry relationship amongst data, while Euclidean distance just computes the pairwise metric between data. Moreover, to achieve faster implementation, the sparse representation in the proposed method is implemented

**Table 4.1:** Notations

| | |
|---|---|
| $m$ | The dimension of feature space |
| $n$ | The number of voxels in the volume |
| $X$ | Features as input data in $\mathbb{R}^{m\times n}$ |
| $k$ | The size of dictionary |
| $D$ | Dictionary in sparse coding in $\mathbb{R}^{m\times k}$ |
| $d_i$ | The $i$th atom in $D$ |
| $\alpha$ | Coefficient matrix in $\mathbb{R}^{k\times n}$ over $D$ |
| $\alpha_i$ | Coefficient vector in $\mathbb{R}^{k\times 1}$ of $i$th data over $D$ |
| $d$ | The dimension of output space |
| $P$ | Mapping in $\mathbb{R}^{d\times m}$ from input to output space |
| $Y$ | Mapping features in $\mathbb{R}^{d\times n}$ |
| $\lambda$ | Regularization parameter for the sparsity of $\alpha$ |
| $\lambda_G$ | Regularization parameter for the geometrical constraint |
| $K$ | Number of neighbors in the neighbor graph |
| $KK$ | Number of neighbors when applying Isomap and LLE to $D$ |
| $\| \bullet \|_F$ | Frobenius norm |
| $Vol(C)$ | Volume of the parallelotope spanned by the clique $C$ |
| $G(C)$ | Gramian matrix |

based on the stochastic gradient descent Bottou and Bousquet (2008); Mairal et al. (2009), which minimizes the sequential approximation of the objective function. For simplicity, the notations employed in this paper are tabulated in Table 4.1.

## 4.2 Overview of Geometry Constrained Sparse Embedding

As aforementioned, the transfer function design in DVR is reduced to the dictionary learning. To obtain a better dictionary, the intrinsically geometric structure of feature vectors are taken into account in GCSE in this paper.

As shown in Fig. 4.2, for each datum in a high dimensional data set, a group of neighbors are selected according to a geometry distance of the datum to the hyperplane in the high dimensional manifold, and a neighbor graph is constructed

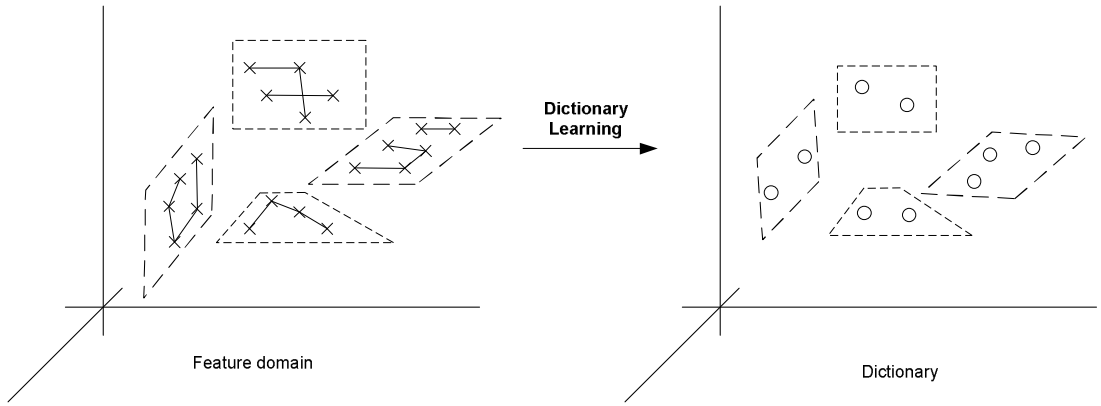



**Figure 4.2:** Illustration of dictionary learning.

to encode the geometrical information into GCSE. The learnt dictionary can better describe the features with the similar pattern, so that such dictionary can act as a surrogate for the entire features. The framework of GCSE is summarized in Algorithm (2). Steps 2 and 3 are to be discussed in details in following sections.

**Algorithm 2** Overview of GCSE.

**Input:** CT data**,** number of neighbors $K$, size of dictionary $k$**,** initial dictionary $D_0 \in \mathbb{R}^{m\times k}$**.**

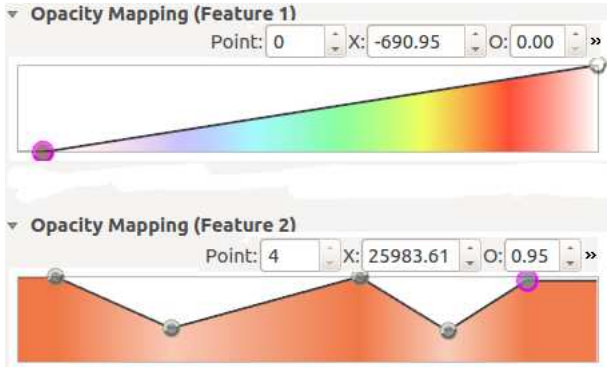
1: Extract features of CT data $X \in \mathbb{R}^{m\times n}$.
2: Construct a neighbor graph using the geometry distance.
3: Learn the dictionary $D^*$ based on the objective function encoding the geometrical structure.
4: Apply the conventional DR algorithm to $D^*$ to find the mapping $P \in \mathbb{R}^{d\times m}$.

**Output:** $Y = PX \in \mathbb{R}^{d\times n}$**.**

Once feature vectors $Y = [y_i]_{i=1}^d$ is generated using the proposed GCSE method, a multi-dimensional transfer function is designed to assign the color and opacity, $Y \to (r, g, b, \alpha)$. In this paper, the transfer function is design in a simple way. As shown in Fig. 4.3, the color $(r, g, b)$ is mapped based on $y_1$, and the opacity is assigned using the production of different opacity mappings, which are generated based on all features $Y$.

$$\alpha = \Pi_{i=1}^d \alpha(y_i) \tag{4.1}$$

**(a)** Color mapping




**(b)** Opacity mapping.

**Figure 4.3:** Design of multi-dimensional transfer function.

## 4.3 Neighbor Selection Using Geometry Distance

A neighbor graph is constructed to describe the locally geometric relationship of the feature vectors. As a measure, Euclidean has shown good performance in the neighbor selection. Nevertheless, Euclidean distance is limited as only the pairwise distance to the target data is considered. It is desired that the local relationships in the cluster are also taken into account. In this section, a geometry distance is computed to measure the geometrical similarity of the features.

### 4.3.1 Geometry Distance

The fundamental problem in construction of neighbor graph is how to calculate the distance during the neighbor selection. In this section, a distance related to geometric information of input features is presented with the assumption that each data and the neighbors lie on a linearly local patch. In other words, the data can be reconstructed linearly by the neighbors in that patch. By defining this geometry

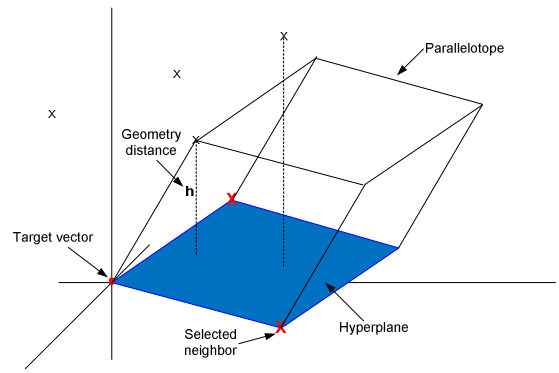



**Figure 4.4:** Illustration of geometry distance in three dimensions.

distance as the neighbor criterion, we can explore and preserve the geometry relation of input features.

The difference from Euclidean distance is that the geometry one is expressed by defining the distance from a feature vector to a clique, which consists of previously selected neighbors. To facilitate calculation of this distance, all feature vectors are translated into the coordinate with the origin at the target vector. As shown in Fig. 4.4, the red dot denotes the target vector, and cross makers in red represent the selected neighbors, which form a hyperplane. The geometry distance $h$ is the distance from a new coming feature vector $c_K$ to the hyperplane. To calculate $h$, a parallelotope is built by $C = [c_i]_{i=1}^{K} \in \mathbb{R}^{m \times K}$ including the selected neighbors $[c_i]_{i=1}^{K-1}$ and $c_K$. The volume of this parallelotope is firstly computed Meyer (2000) by

$$\begin{aligned}
\mathrm{Vol}(C) = &\ \det{}^{\frac{1}{2}}(C^T C) \\
= &\ \det{}^{\frac{1}{2}}\left(\begin{bmatrix} c_1^T \\ c_2^T \\ \vdots \\ c_K^T \end{bmatrix} \begin{bmatrix} c_1^T & c_2^T & \cdots & c_K^T \end{bmatrix}\right) \\
= &\ \det{}^{\frac{1}{2}}\begin{pmatrix} c_1^T c_1 & \cdots & c_1^T c_K \\ \vdots & \vdots & \vdots \\ c_K^T c_1 & \cdots & c_K^T c_K \end{pmatrix} = \det{}^{\frac{1}{2}}(G(C))
\end{aligned} \tag{4.2}$$

following the definition of Gramian Barth (1999). We further illustrate the distance from a new feature vector to the selected clique and judge whether a set of vectors are linearly independent in the space. The distance from $c_K$ to the hyperplane spanned by $C_{[1,(K-1)]} = [c_i]_{i=1}^{K-1}$ can be represented Meyer (2000); Barth (1999) with the volume of the parallelotope

$$h^K_{[1,(K-1)]} = \sqrt{\frac{\mathrm{Vol}^2(C_{[1,K]})}{\mathrm{Vol}^2(C_{[1,(K-1]})}}. \tag{4.3}$$

Following the properties of block matrices, it can be further obtained that

$$\begin{aligned} & \mathrm{Vol}^2(C_{[1,K]}) \\ = \; & \det([C_{[1,(K-1)]}, c_K]^T [C_{[1,(K-1)]}, c_K]) \\ = \; & \det(G(C_{[1,(K-1)]})) \det(c_K^T c_K - (c_K^T C_{[1,(K-1)]}) \times \\ & G^{-1}(C_{[1,(K-1)]})(C^T_{[1,(K-1)]} c_K)) \\ = \; & \det(G(C_{[1,(K-1)]})) \det(G'). \end{aligned} \tag{4.4}$$

Note that $G' \in \mathbb{R}^{1\times 1}$,

$$\begin{aligned} G' = \; & \det(c_K^T c_K - (c_K^T C_{[1,(K-1)]}) \times \\ & G^{-1}(C_{[1,(K-1)]})(C^T_{[1,(K-1)]} c_K)) \end{aligned} \tag{4.5}$$

and according to (4.2),

$$\mathrm{Vol}^2(C_{[1,(K-1)]}) = \det(G(C_{[1,(K-1)]})), \tag{4.6}$$

thus Eq. (4.3) can be rewritten as

$$h^K_{[1,(K-1)]} = \sqrt{\det(G')} = \sqrt{G'}. \tag{4.7}$$

Using the above equation, for each target data, a set of geometry distances are calculated and sorted, the neighbor can be found by selecting one corresponding to the shortest distance.

### 4.3.2 Implementation of Neighbor Selection

In the implementation of neighbor selection, two issues have to be considered. The first one is the computational burden. For the feature vectors extracted from the huge-size medical data, neighbor selection is a time-consuming procedure using geometry distance, as well as Euclidean distance, because all feature vectors are taken into account to find a right neighbor. Second, linear dependence of selected neighbors in the clique may cause the volume of parallelotope $\mathrm{Vol}^2(C_{[1,(K-1]})$ to be zero. In other words, inverse matrix of $G(C_{[1,(K-1)]})$ in $G'$ doesn't exist. More details in regard of both issues are discussed in the following.

#### Selection of neighbor candidates

Neighbor candidates are chosen with the assumption that neighbors are usually distributed around the target vector. Therefore, a sphere is extracted according to the index of voxel in the medical data, and the corresponding feature vectors form the neighbor candidates. In this paper, the radius of sphere $r$ is determined by the desired number of neighbors $K$. In the experiment, $r$ is set to $2K$ empirically. Much less feature vectors are calculated in the following neighbor selection. This greatly saves the computational resource without significant loss on performance.

#### Neighbor selection

Considering the linear independence, the geometry distance can be applied only when the number of neighbors $K$ is less than or equal to the dimension of the dimension of feature vector $m$. That is, the volume is defined with respect to the dimension of the feature vector. For instance, a non-zero volume of a polyhedron may become zero when it is considered in a higher dimensional space. However, $K$ is usually greater than $m$ in practice, that leads to the unavoidable dependence. To address this problem, the neighbor selection can be simplified without much loss on performance. Actually there is no need to compute geometry distances for all selected neighbors,

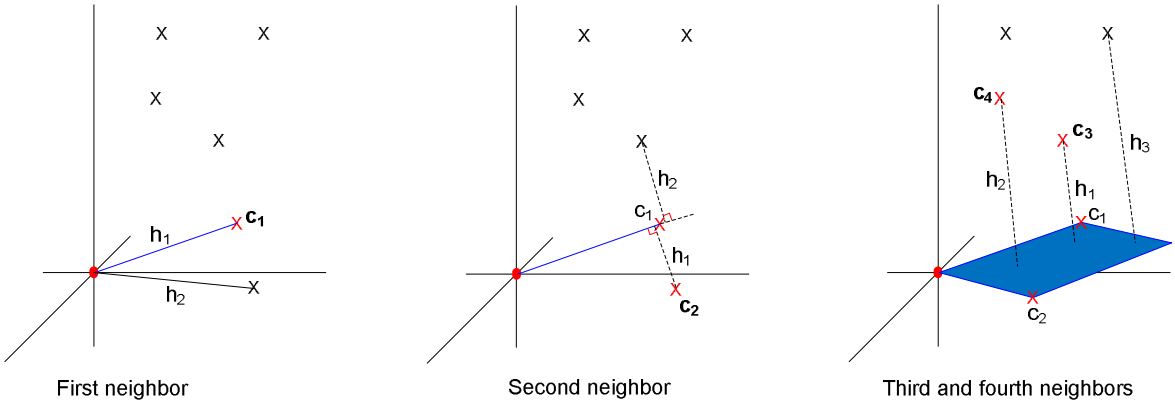



**Figure 4.5:** Example of neighbor selection using geometry distance.

especially when the number of selected neighbors exceeds the dimension of feature vector. It is redundant to calculate (4.7) for following neighbors if a parallelotope has been approximately determined by selected neighbors. Therefore, it is possible to cease increasing the number of sides of a parallelotope once the selected number of neighbors is equal to the dimension of feature vector.

An alternative method for neighbor selection is to find the hyperplane determined by the kernel vectors. After a hyperplane is obtained, other neighbors are chosen according to the distance to this hyperplane. Given a set of vector data, the maximal number of kernel vectors is the dimension of the vector space, provided that any vectors in the space can be linearly represented by the kernel vectors. A hyperplane with dimension $d-1$ can be spanned by $d$ linearly independent kernel vectors. To have an intuitive view, an example is shown in Fig. 4.5. The dimension of feature vector is 3, and the number of required neighbors is set to 4. The red dot is the target vector. When the target vector is determined, the first neighbor $c_1$ is obtained according to the distance to the target. This is in common with the selection using Euclidean distance. The second one $c_2$ is chosen according to the distance from the new feature vector to the line through the target and first neighbor. At this point, a parallelogram with the dimension 2 is obtained with the target and selected neighbors

as vertices. The third and fourth neighbors are determined by choosing the first two shortest distances to this parallelogram.

Furthermore, even though the number of neighbors are less than the dimension of feature vector, the linear dependence probably happens. Hence, the inverse matrix of $G(C_{[1,(K-1)]})$ does not exist. A simple technique to solve this problem is employing the Tikhonov regularization

$$\begin{aligned} G' = \quad & \det(c_K^T c_K - (c_K^T C_{[1,(K-1)]}) \times \\ & (G(C_{[1,(K-1)]}) + \epsilon I)^{-1} (C_{[1,(K-1)]}^T c_K)) \end{aligned} \tag{4.8}$$

with $\epsilon > 0$ as the regulation parameter.

## 4.4 Dictionary Learning

In this section, the dictionary is learnt by encoding the geometric relation of feature vectors, and the stochastic gradient descent method is employed to speed up the implementation of dictionary learning.

### 4.4.1 Objective Function

Let $D = [d_1, d_2 \cdots d_k] \in \mathbb{R}^{m \times k}$ be the dictionary matrix, and $\alpha = [\alpha_1, \alpha_2 \cdots \alpha_n] \in \mathbb{R}^{k \times n}$ denotes the sparse matrix. Given a data set $X = [x_1, x_2, \cdots x_n] \in \mathbb{R}^{m \times n}$, the objective function of original sparse representation can be defined as

$$\begin{aligned} & \{D, \alpha\} = \underset{\{D,\alpha\}}{\arg\min} \tfrac{1}{2} \parallel X - D\alpha \parallel_F^2 + \lambda \parallel \alpha \parallel_1 \\ & s.t. \parallel d_i \parallel^2 \leq 1, \ i = 1, \cdots k, \end{aligned} \tag{4.9}$$

where $\lambda$ is the regularization parameter. To prevent arbitrarily large values of $D$, the constraint $\parallel d_i \parallel^2 \leq 1$ is applied. Base on this basic objective function, an extra item is integrated to preserve the geometrical structure of $X$. To achieve it, we can construct a neighbor graph $G$ with $n$ vertices according to the geometry distance in Section 4.3.

Each vertex represents a feature vector of medical data, and it is associated with a element in the weight matrix $W$ of $G$. If $x_j$ belongs to the neighbors of $x_i$, $w_{i,j} = 1$, otherwise, $w_{i,j} = 0$. The difference with the graph constructed by Euclidean distance is that the weight matrix $W$ is an asymmetric matrix. That is, if $x_i$ is a neighbor of $x_j$, $x_j$ may not be the neighbor of $x_i$.

To map the weighted graph $G$ to the sparse coefficients $\alpha$, a reasonable criterion Belkin and Niyogi (2001) is chosen by minimizing the following function:

$$\begin{aligned} & \tfrac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \parallel \alpha_i - \alpha_j \parallel_F^2 w_{i,j} \\ = \; & \tfrac{1}{2}(\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\alpha_i^T w_{i,j} + \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_j\alpha_j^T w_{i,j} \\ & -2\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i^T\alpha_j w_{i,j}) \\ = \; & \tfrac{1}{2}(\mathbf{tr}(\alpha R\alpha^T) + \mathbf{tr}(\alpha C\alpha^T)) - \mathbf{tr}(\alpha W\alpha^T) \\ = \; & \mathbf{tr}(\alpha L\alpha^T), \end{aligned} \tag{4.10}$$

where $L = \frac{1}{2}(R + C) - W$ is the Laplacian matrix. The degrees of $x_i$ along the row and column are defined as $R$ and $C$

$$\begin{aligned} R = diag(r_1, r_2, \cdots r_N),\ r_i = \textstyle\sum_{j=1}^{N} w_{i,j} \\ C = diag(c_1, c_2, \cdots c_N),\ c_j = \textstyle\sum_{i=1}^{N} w_{i,j}. \end{aligned} \tag{4.11}$$

By combining Eq. (4.9) and Eq. (4.10), we can obtain following objective function:

$$\begin{aligned} & D^* = \arg\min_{D} \tfrac{1}{2} \parallel X - D\alpha \parallel_F^2 + \lambda \parallel \alpha \parallel_1 + \lambda_G tr(\alpha L\alpha^T) \\ & s.t. \parallel d_i \parallel^2 \leq 1,\ i = 1, \cdots k, \end{aligned} \tag{4.12}$$

where $\lambda_G$ is the regularization parameter for the geometric structure.

In this paper, stochastic gradient descent is employed to handle the large amount of features of medical data. The dictionary learning algorithm is summarized in Algorithm 3. The dictionary and coefficients can be optimized alternatively by minimizing over one while keeping the other one fixed Lee et al. (2007). To improve

convergence rate further, $\tau$ feature vectors are selected randomly each iteration. Then the sparse coefficient $\alpha_{t,i}$ of $x_{t,i}$ over the previous dictionary $D_{t-1}$ is computed by minimizing Eq. (4.13). $A$ and $B$ are calculated to update the dictionary using Algorithm 4. The details are discussed in Section 4.4.3.

**Algorithm 3** Dictionary Learning.

**Input:** $X \in \mathbb{R}^{m\times n}$, $\lambda, \lambda_G, \lambda_L \in \mathbb{R}$, initial dictionary $D_0 \in \mathbb{R}^{m\times k}$, batch size $\tau$, number of iterations $T$.

1: $A_0 \leftarrow 0$, $B_0 \leftarrow 0$.
2: **for** $t = 1$ to $T$ **do**
3: Select $\tau$ feature vectors $\{x_{t,i}\}_{i=1}^{\tau}$ randomly.
4: Perform sparse coding using feature-sign search algorithm for each vector $x_{t,i}$.

$$\begin{aligned}\alpha_{t,i} = \arg\min_{\alpha\in R^{k\times 1}} \tfrac{1}{2} \parallel x_{t,i} - D_{t-1}\alpha \parallel_F^2 + \lambda \parallel \alpha \parallel_1 \\ + \lambda_G \tfrac{1}{2} \sum_{i=1}^{n}\sum_{j=1}^{n} \parallel \alpha_i - \alpha_j \parallel_F^2 w_{i,j}.\end{aligned} \tag{4.13}$$

5: $A_t \leftarrow A_{t-1} + \sum_{i=1}^{\tau}(\alpha_{t,i}\alpha_{t,i}^T)$.
6: $B_t \leftarrow B_{t-1} + \sum_{i=1}^{\tau}(x_{t,i}\alpha_{t,i}^T)$.
7: Compute $D_t$ based on block-coordinate descent, refer to Section 4.4.3 for details.
8: **end for**

**Output:** $D^*$.

### 4.4.2 Sparse Coding

In this section, the sparse coding problem is solved by fixing the dictionary $D$. According to Algorithm 3, each sparse coefficient vector $\alpha_{t,i}$ is updated individually, while the rest are fixed. In order to optimize $\alpha_{t,i}$ individually, Eq. (4.12) should be formulated in a vector form. The combination of residual item and sparse regularizer can be denoted as $\sum_{i=1}^{n} \parallel x_i - D\alpha_i \parallel_F^2 + \lambda \sum_{i=1}^{n} \parallel \alpha_i \parallel_1$, where the subscript $t$ from $x$ and $\alpha$ is omitted for simplicity. The Laplacian regularizer $\mathbf{tr}(\alpha L \alpha^T)$ can be expanded as follows:

$$\begin{aligned}\mathbf{tr}(\alpha L \alpha^T) &= \mathbf{tr}\left(\sum_{i,j=1}^{n} L_{i,j}\alpha_i\alpha_j^T\right) \\ &= \sum_{i,j=1}^{n} L_{i,j}\alpha_i^T\alpha_j.\end{aligned} \tag{4.14}$$

Thus, the problem Eq. (4.12) can be rewritten as

$$\begin{aligned}\alpha^* &= \arg\min_{\alpha} \tfrac{1}{2}\sum_{i=1}^{n} \parallel x_i - D\alpha_i \parallel_F^2 + \lambda \sum_{i=1}^{n} \parallel \alpha_i \parallel_1 \\ &+\lambda_G \sum_{i,j=1}^{n} L_{i,j}\alpha_i^T\alpha_j.\end{aligned} \tag{4.15}$$

Because only $\alpha_i$ is updated one time, by fixing other sparse coefficients, we can obtain the objective function for $\alpha_i$ as follows:

$$\begin{aligned}\alpha_i^* &= \arg\min_{\alpha_i} \tfrac{1}{2} \parallel x_i - D\alpha_i \parallel_F^2 + \lambda \parallel \alpha_i \parallel_1 \\ &+\lambda_G \left(L_{i,i}\alpha_i^T\alpha_i + 2\alpha_i^T \sum_{j\neq i,}^{n} L_{i,j}\alpha_j\right).\end{aligned} \tag{4.16}$$

In this paper, the above problem is solved using the feature-sign search algorithmLee et al. (2007) as a unconstrained quadratic optimization problem. It outperforms many existing algorithms, such as grafting Perkins and Theiler (2003) and LARS Efron et al. (2004).

### 4.4.3 Dictionary Update

In this section, the dictionary update is performed using block-coordinate descent by minimizing residual error in Eq. (4.12). Each column of dictionary is updated sequentially according to Eq. (4.18) under the constraint $\parallel d_i \parallel^2 \leq 1$. $Da_j - b_j$ is the gradient of $f(D) = \frac{1}{2} \parallel X - D\alpha \parallel_F^2$ with respect to $d_j$, and $\eta$ is the learning rate. As mentioned in Section 4.4.1, the dictionary learning is implemented based on stochastic gradient descent algorithm. Taking partial differentiation with respect to $D$, we can obtain

$$
\begin{aligned}
\frac{\partial f}{\partial D} &= -(x_i - D\alpha_i)\alpha_i^T \\
&= D(\alpha\alpha^T) - x_i\alpha_i^T \\
&= DA - B,
\end{aligned} \tag{4.17}
$$

where $A = \alpha\alpha^T$ and $B = x_i\alpha_i^T$. Since the previous dictionary $D_{t-1}$ is used in the proposed algorithm to provide a prior knowledge for computing $D_t$, number of iteration $T$ can be set to a small value. In our experiments, only one iteration is enough for the convergence.

**Algorithm 4** Dictionary update.

**Input:** $D = [d_1, d_2, \cdots d_k] \in \mathbb{R}^{m\times k}$, $A = [a_1, a_2, \cdots a_k \in \mathbb{R}^{k\times k}] = \sum_{i=1}^{t} \alpha_i\alpha_i^t$, $B = [b_1, b_2, \cdots b_k \in \mathbb{R}^{m\times k}] = \sum_{i=1}^{t} x_i\alpha_i^t$, number of iteration $T$.

1: **for** $i = 1$ to $T$ **do**
2: **for** $j = 1$ to $k$ **do**
3: Update the $j$th column of the dictionary:

$$
\begin{aligned}
d_j &\leftarrow d_j - \eta(Da_j - b_j) \\
d_j &\leftarrow \frac{1}{max(\|d_j\|_F, 1)} d_j.
\end{aligned} \tag{4.18}
$$

4: **end for**
5: **end for**

**Output:** $D^*$(Updated dictionary).

## 4.5 Experimental Results and Discussions

In this section, three sets of experiments were performed to evaluate the performance of GCSE on both synthetic and real CT data sets. The GCSE firstly employed the synthetic data to compare with the existing methods, including linear and non-linear approaches. The convergence and the robustness of related parameters were also studied. The second experiment was to test the capability of handling large data set using a partial CT volume extracted from the golden standard data set Pawiro et al. (2011). For the effect of volume rendering using different transfer functions, there is no quantitative way to measure that. Thus, we presented the rendered

**Table 4.2:** Specifications of synthetic and real CT data sets.

| Data source | Size of data set | Size of features | Target dimension |
|---|---|---|---|
| Toroidal helix | $3 \times 800$ | $3 \times 800$ | 2 |
| Gold data | $20 \times 21 \times 20$ | $6 \times 8400$ | 2 |
| Human tooth | $94 \times 81 \times 155$ | $6 \times 1180170$ | 2 |
| Mixed phantom | $395 \times 258 \times 201$ | $6 \times 20483910$ | 2 |

images that successfully emphasize the region of interests to compare the difference for visualization. Human tooth and mixed phantom CT data were used in the last test. Particularly, the mixed phantom consists of a spinal phantom, empty bottle, tape measure and so forth. The corresponding CT data is collected by a Aquilion ONE scanner (Toshiba, Japan) with a 320-row detector.

The specifications of data sets used in the experiments are summarized in Table 4.2. Toroidal Helix data set is used directly as the input of dimensionality reduction methods. For the real CT data sets, corresponding feature vectors are both in 6 dimensions. The first vector is related to the intensity for each voxel. To discriminate the voxels next to the boundaries. the following 4 ones are the gradient information, including gradient magnitude and three components in the gradient vector. The last one is the Hessian measure, which can capture the edge information to deal with complex configurations.

### 4.5.1 Comparison with Conventional Methods

In this experiment, Toroidal Helix data set was used to compared GCSE with conventional methods, including PCA, Isomap and LLE. "Toroidal Helix" is generated by coiling an one dimensional curve around a helix. As shown in Fig. 4.6, all conventional methods project the input data into a symmetric and flower-like shape, and preserve the structure and distribution of the helix. In this test, the number of neighbors are 40 and 60 for Isomap and LLE, respectively, while PCA does not need any parameters for the dimensionality reduction.

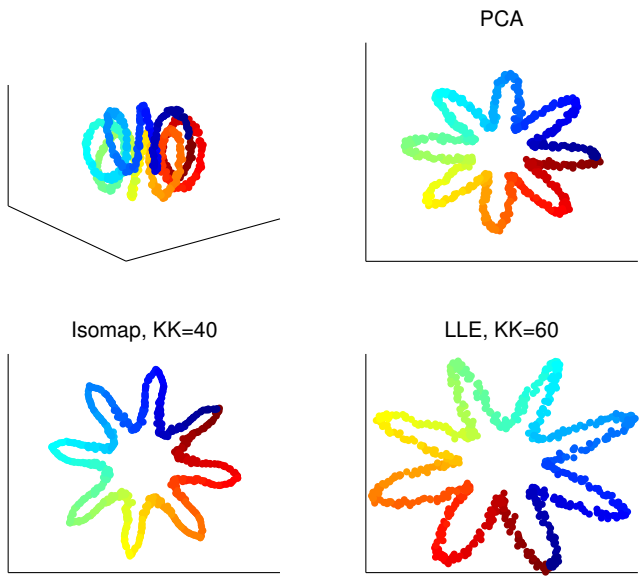



**Figure 4.6:** Results of conventional methods.

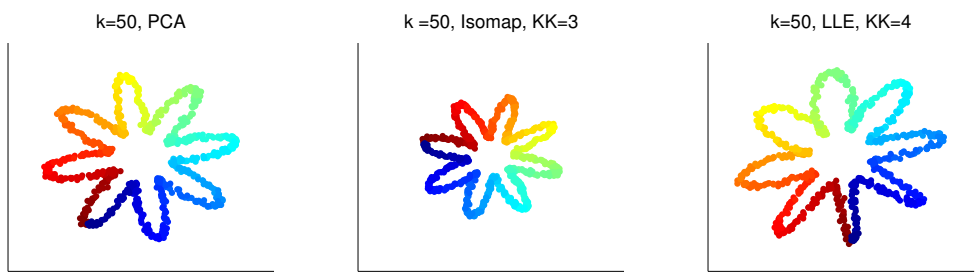



**Figure 4.7:** Results of GCSE.

Fig. 4.7 shows the results of GCSE on Toroidal Helix data set. When the learnt dictionary $D$ is ready, we apply PCA, Isomap and LLE to $D$, respectively. In this experiment, the size of dictionary is 50, the numbers of neighbors for Isomap and LLE are 3 and 4. We can see that much less neighbors are required to achieve the similar effect compared with the conventional methods. It also preserves the geometric information of the input.

To explain the less number of neighbors in GCSE, Fig. 4.8 illustrates the distribution of Toroidal Helix data set and dictionary used for dimensionality reduction. The dictionary is better able to keep the topological structure with the less

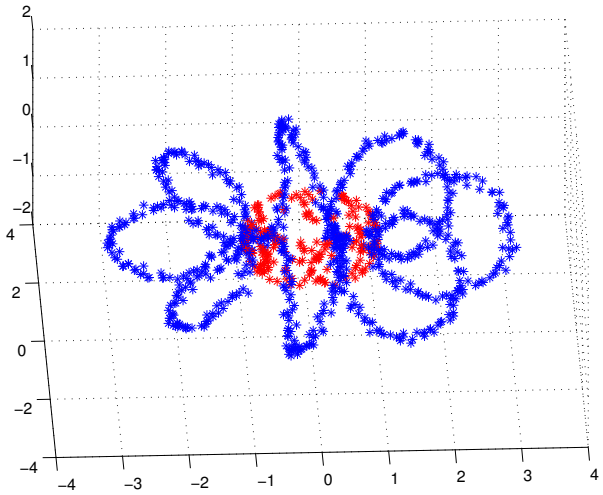


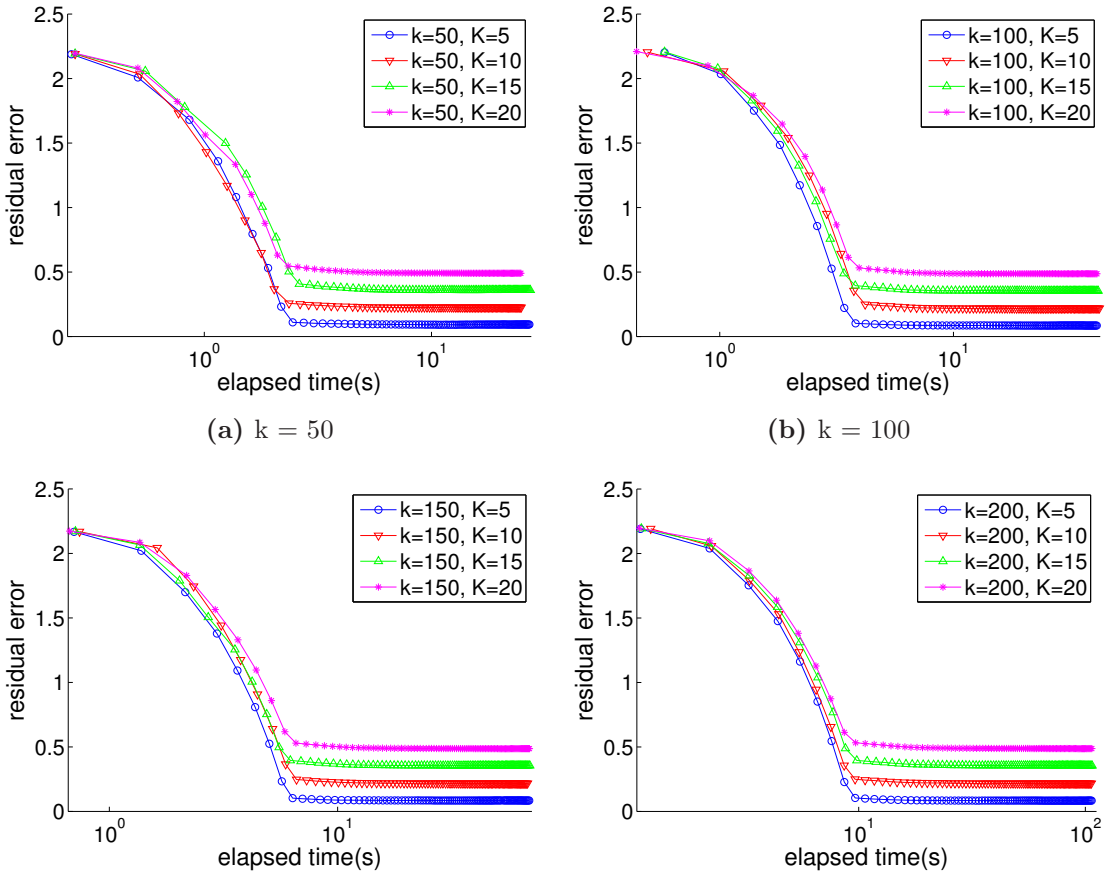

**Figure 4.8:** Topologies of input data and dictionary in 3D. The blue stars are input data, and the learnt dictionary is indicated by the red stars.

atoms, so that the local structure can be presented using less points in the dictionary, while more neighbors are needed to express the same geometric information.

We also studied the convergence and robustness of GCSE based on the stochastic gradient descent algorithm. Following the convergence analysis of stochastic gradient descent in Shapiro and Wardi (1996), it is obvious that the dictionary can approximate to a stationary point asymptotically. To verify this, we set different sizes of dictionary and numbers of neighbors indicating the geometric structure in the objective function. In the test, the sizes of dictionary were $\{50, 100, 150, 200\}$, and the values for the number of neighbors were $\{5, 10, 15, 20\}$. GCSE is performed with the iteration number of 100, and the batch size in each iteration is also set to 100. The residual errors are shown in Fig. 4.9. In all cases, the average residual error is 0.29.

### 4.5.2 Comparison with Sparse Embedding

This experiment was implemented to test the capability of large data set handling using Nguyen's SE and GCSE. We tried to do the dimensionality reduction on gold



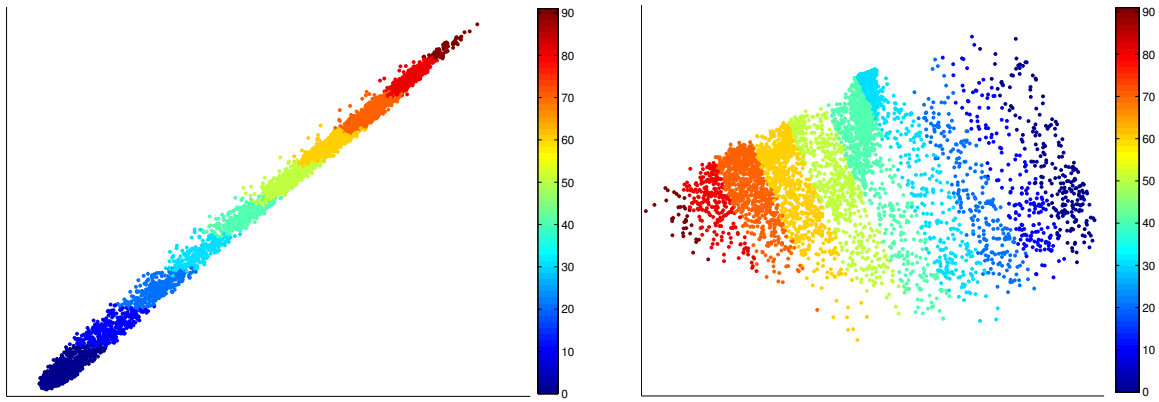

**(a)** k = 50 **(b)** k = 100

**(c)** k = 150 **(d)** k = 200

**Figure 4.9:** Residual errors using different configurations on Toroidal Helix data set.

standard data set and human tooth CT. However, SE failed using both data sets due to the memory limit. The same problem occurred when we apply the conventional methods to these data sets. GCSE can deal with this case, the evaluation of GCSE on the large data set is presented in next section.

To compare the performance of GCSE and SE, a partial CT volume with the size of $20 \times 21 \times 20$, extracted from gold standard data, was used. The feature vectors are in 6D, and the target dimension is 2. In SE, the iteration number was set to 5, and A polynomial kernel of degree 4 was used. For GCSE, the size of target dictionary was set to 200, the iteration number was 100, and the size of batch processing in




**(a)** SE. **(b)** GCSE.

**Figure 4.10:** Comparison between sparse embedding and proposed method.

each iteration was 400. LLE was used to perform the dimensionality reduction on the dictionary.

In order to facilitate the visualization of feature vectors in the reduced dimension, the regions were assigned with different colors according to the intensity information. Fig. 4.10 shows the results of SE and GCSE. Both of them can discriminate the regions correctly. The difference is that GCSE preserves the additional geometric information in the original data, while SE just embedded the features into different areas. For the computing time, GCSE takes around 3.74 s each iteration, and SE requires about 862.93 s for every round. Although the number of iteration for SE is less than GCSE's, SE is time-consuming when the total computing time is taken into account.

### 4.5.3 Evaluation of Transfer Function in DVR

In this section, the rendering quality of boundaries in the direct volume rendering is mainly evaluated using different transfer functions on human tooth and mixed phantom. Conventional methods and SE can not handle these data sets with the size. In this experiment, two transfer functions are designed for comparison. One transfer

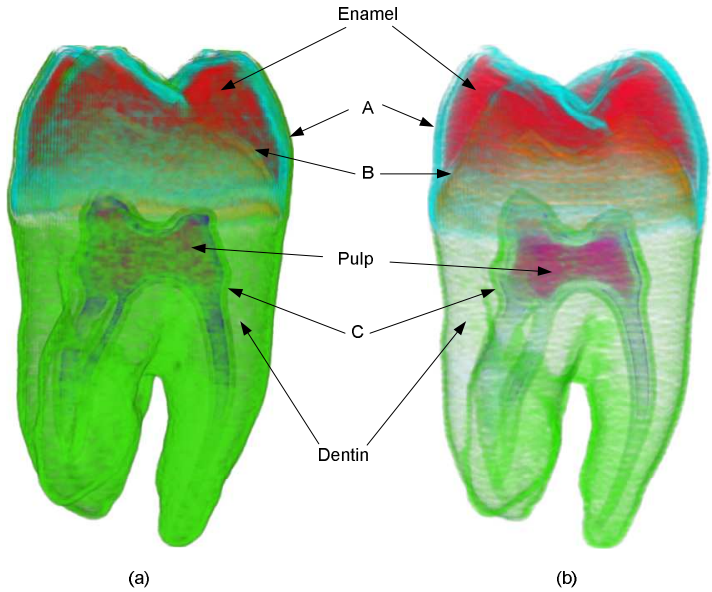



**Figure 4.11:** Rendered images using different transfer functions. (a) is the result using the transfer function based on the intensity and gradient information. (b) shows the rendered image based on the features obtained by GCSE. A, B and C indicate the background/enamel, pulp/dentin, background/dentin boundaries respectively.

function (TF-IG) was designed based on the intensity and gradient information, the other one (TF-GCSE) was obtained using the proposed GCSE method.

The rendering results are shown in Fig. 4.11. One of the disadvantages for DVR using TF-IG is that some regions close to the boundary can not be isolated exactly. As shown in Fig. 4.11a, when the background/dentin boundary is labeled with the desired color and opacity, the pulp/dentin and background/enamel boundaries are also colored. Because they belong to the same range defined by the intensity and gradient. In Fig. 4.11b, TF-GCSE provides a clearer view on the boundaries. Because the Hessian measures are involved in the resultant features obtained by GCSE. Materials involved in the human tooth can also be identified based on TF-GCSE, such as enamel, dentin and pulp.

We also evaluate the proposed method on mixed phantom. Fig. 4.12 shows the rendered result based on intensity and gradient. The bottle and coat of tape measure are labeled with the same color. That is, we can not discriminate both parts only

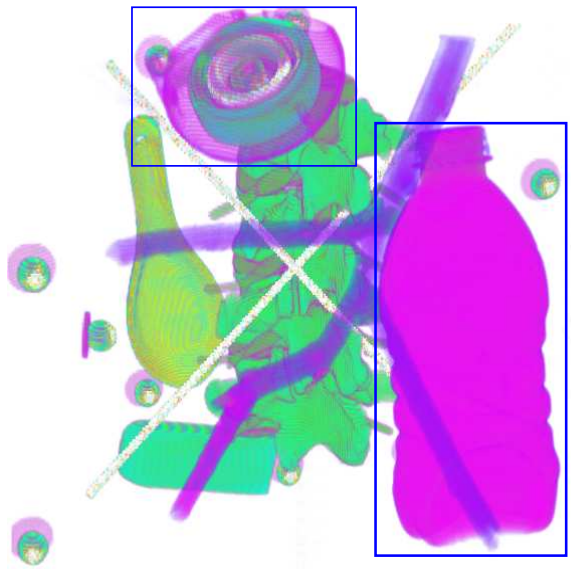


**Figure 4.12:** Rendered image of spinal phantom based on intensity and gradient.

on intensity and gradient. Using the proposed method, 6-D features were mapped in 2D. A transfer function TF-GCSE was designed based on new 2D features, the rendered result is shown in Fig. 4.13. The color was assigned based on feature 1, and the opacity is mapped by combining both opacity settings on feature 1 and feature 2. The color and opacity can be adjusted to separate the bottle only.

## 4.6 Related Work

In this paper, GCSE integrates three different disciplines: multi-dimensional transfer function for direct volume rendering, dimensionality reduction and sparse representation for machine learning. In the following, these fields related to our method are briefly discussed.

### 4.6.1 Multi-dimensional Transfer Function

Transfer function design is still a challenge in direct volume rendering due to its difficulty of design Pfister et al. (2001), especially when the domain of transfer function

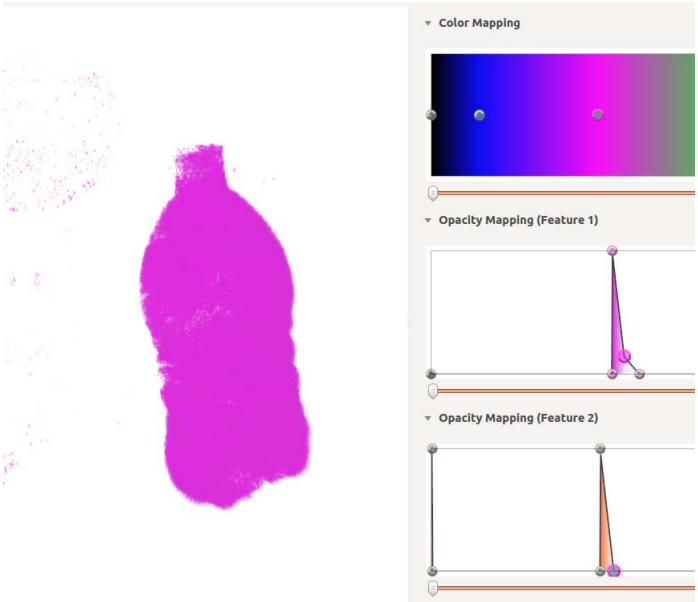



**Figure 4.13:** Rendered image of spinal phantom based on proposed method.

is extended from one-dimensional transfer functions based on the intensity per voxel to multi-dimensional ones. Some multi-dimensional transfer functions have been proposed based on the gradient magnitude Kindlmann and Durkin (1998); Kniss et al. (2002) and curvature information Hladuvka et al. (2000); Kindlmann et al. (2003). The former feature captures the information on boundaries of the volume, and it has become a significant choice for the second dimension in transfer functions. The curvature is used to emphasize the ridges and valleys in the volume. It may enhance the visualization of boundaries further by combining the features of gradient and curvature. Sereda *et al.* Sereda et al. (2006) enhanced the visualization on boundaries using a method based on LH histograms in which a path across the boundary integrating the gradient field was generated. To facilitate the interaction with transfer functions, Kniss *et al.* Kniss et al. (2002) also developed GUI widgets and dual-domain operations. In this paper, we use Kniss's widgets to visualize the volumetric data based on features in the reduced domain.

Besides, several data-centric methods were proposed. In Correa and Ma (2008), a size-based transfer function was proposed by classifying the features based on the relative size in 3D scale field. Caban *et al.* Caban and Rheingans (2008) designed transfer functions for direct volume rendering based on the texture properties, which are captured by combining the first-, second-, and high-order statistics, such as mean, variance and skewness. In Praßni et al. (2010), a shape-based transfer function for the volume visualization was proposed based on the rough pre-segmentation. Longitudinal, surface-like and blobby shapes are distinguished using a shape descriptor.

### 4.6.2 Dimensionality Reduction and Sparse Representation

Due to the limited perception of 4D or higher dimensional data, the dimension of transfer function has to be kept in 2D or 3D for the pratical interaction. Dimensionality reduction is a solution for the design of multi-dimensional transfer functions. The detailed discussion on dimensionality reduction can be found in Van der Maaten et al. (2009). Several dimensionality reduction methods have been applied, including the linear and non-linear approaches. As a linear method, PCA has been successfully applied in many fields. In Salama et al. (2006), a new level of abstraction for transfer function models was introduced based on PCA to facilitate the assignment of optical properties in direct volume rendering. Takanashi *et al.* Takanashi et al. (2002) reduced the dimension of transfer function domain using ICA.

For applications using non-linear methods, Francisco *et al.*de Moura Pinto and Freitas (2007) simplified the transfer function design by the dimensionality reduction using self-organizing maps. In Kim et al. (2010), two non-linear methods, Isomap and LLE, were applied to reduce the dimension of transfer function to a manageable size. Zhao *et al.* Zhao and Kaufman (2010) used LLE to reduce the computational burden and represent the transfer function concisely.

Recently sparse representation has attracted a growing interest, and various applications are implemented in many fields, such as face recognition (FR), dense error correction, motion recognition Wright et al. (2010). In Zhang et al. (2010), the sparse representation based classification for FR was discussed based on the dimensionality reduction. In Huang et al. (2011), a novel algorithm called Sparse Reconstruction Embedding(SRE) was proposed, including two steps, sparse reconstruction and low-dimensional embedding. The concept of later step is similar to the final step of LLE. A sparse embedding framework for dimensionality reduction was presented in Nguyen et al. (2012). The dictionary was learnt simultaneously in the reduced dimension by combining sparse representation and dimensionality reduction.

## 4.7 Summary

In this paper, a novel scheme for dimensionality reduction is proposed by encoding the geometrical structure based on a geometry distance. GCSE performs well by learning a good dictionary as a problem of SR. To the best of our knowledge, it is the first time to introduce SR into the design of transfer function in direct volume rendering. More importantly, GCSE enhances the capability to deal with the large data set and enables fast implementation based on stochastic gradient descent. Experimental results show the convergence and robustness of GCSE on the synthetic data. We also have tested GCSE on the CT data containing millions of samples, and the features in the reduced dimension are able to discriminate the materials and boundaries in the volume.

# Chapter 5

# Conclusions and Future Work

In this dissertation a mobile C-arm positioning system is developed for image-guided surgery to reduce the radiation exposure to the patient and surgeons, reduce the duration of operation, and enhance the capability of visualization. This system integrates the pre-operative volume, intra-operative images and 3D surface model of patient. It allows surgeons to visualize the hidden anatomy directly and identify the pathology from the outside. More importantly, it can be integrated into the current workflow seamlessly, and also applicable to the training, planning and diagnosis phases of treatment where appropriate treatment options are developed and refined to preserve as much native tissue as possible without compromising surgical goals.

## 5.1 Conclusions

A framework of mobile C-arm positioning system is proposed with the aid of virtual reality. In order to reduce the unnecessary radiation exposure to the patient and surgeons and improve the efficiency of C-arm positioning, two aspects are taken into account, including the determination of reference pose for mobile C-arms and control of mobile C-arms. On one hand, the reference pose is figured out with respect to the outside of patient based on VR in which the inside pre-operative volume and outside surface model of patient are aligned. In this research, the surface model of patient

is constructed using a single camera mounted on the mobile C-arm. Along with the rotation of C-arm, several images are recorded, and a dense model is obtained by sequentially employing the techniques of structure from motion and multi-view stereo. A novel registration method is presented to align the inside and outside of patient. Intensity-based 2D/3D registration is used to transform the inside anatomy to the coordinate system of tracking system. The outside model is also registered in a common coordinates by tracking the position of camera. No calibration procedure is required in advance. On the other hand, to control the mobile C-arm to the desired pose, the mobile C-arm is modeled as a robotic arm with a movable base. In order to increase the mobility and flexibility in the congested operating room, three more degrees of freedom assigned to the base of C-arm. Experimental results shows that the proposed positioning system satisfies the clinical requirement, and additional DoFs enable a larger working space than the standard model.

During the alignment of patient's inside and outside model, intensity-based 2D/3D registration based on Digitally Reconstructed radiographs (DRRs) plays an important role. Although DRR-based method provides a high accuracy, the small capture range hinders its clinical use due to the initial guess. To address such problem, a robust and fast initialization method is proposed using a two-level scheme, including automatic tracking based initialization (Level I) and multi-resolution estimation based on central-slice theorem and phase correlation in Fourier domain (Level II). Level I is achieved by the Vicon motion tracking system, and the resultant transformation is used for the initialization of Level II. This hardware-software integrated approach provides almost optimal transformation parameters for intensity-based registration. A public gold standard data set and a spinal phantom are conducted in the experiments. The mean target registration error (mTRE) was less than 22.57 mm after tracking based initialization only. The intensity-based 2D/3D registration using proposed two-level initialization achieved the successful rate of 84.8% with the average error of 2.36 mm. The experimental results showed that the proposed method yielded the robust and fast initialization for intensity-based registration methods.

To help surgeons identify and locate the pathology in the pre-operative volume, high-quality visualization is necessary. Direct volume rendering (DVR) is commonly employed for the medical visualization with the multi-dimensional transfer functions, which are used to emphasize the region of interest in details. The third goal is mainly to address the problem of transfer function design, when the functions are in more than three dimension. That is, it is impractical to interact directly through a computer interface. Considering the huge size of volume data, a novel framework called geometry constrained sparse embedding (GCSE) for dimensionality reduction (DR) is proposed using the concept of sparse representation. The dictionary leant from the sparse representation acts as the surrogate, which preserves the pattern and geometric information of features. The mapping is derived approximately using this dictionary and feeds to the original features to obtain the ones in the reduced dimension. A geometry distance is presented to capture the geometric relation among the features during the dictionary learning based on stochastic gradient descent algorithm. To the best of our knowledge, it is the first time to introduce sparse representation into the design of multi-dimensional transfer function in direct volume rendering.Various experiments have been conducted using both synthetic and real CT data sets. GCSE performed well on the synthetic data, similar results are produced compared with conventional methods. More importantly, GCSE is implemented with the capability to handle the large data set more powerfully, while conventional methods and sparse embedding approach can not do it due to the memory limitation. The rendering results using the features in the reduced dimension have demonstrated the effectiveness of GCSE.

## 5.2 Future Work

The proposed mobile C-arm positioning system can greatly reduce the radiation exposure in the image-guided surgery. It can be integrated into the current clinical workflow, and no extra management is required. The preliminary experimental

results on the samples of CT data and a spinal phantom prove the practicability of proposed technique. Despite of those, more work is needed for further validating the performance of developed system. The future work is listed as follows:

1. In current research, the proposed system has been verified using the CT data, which is better suited for the diagnosis of bone injuries, lung and chest imaging and cancer detection. The details in term of soft tissue is lost, because no single imaging modality provides the optimal definition of both hard and soft tissues. To this end, MRI may be involved for soft tissue (e.g. ligament and tendon, spinal cord, brain tumors) by multi-modal registration of anatomical structures. A synthetic virtual model is constructed to provide the surgeon with the most complete information about the anatomy and physiology.

2. For the visualization purpose, a user interface will be developed to facilitate the interaction, GPU support will be also considered to ensure a highly interactive rendering speed. For large data sets, to avoid the poor-quality image during the interaction, popular acceleration techniques such as empty space skipping, early ray termination, and adaptive sampling may be utilized. This is to ensure that computations that do not contribute to the final rendered image are avoided as early as possible in the rendering pipeline. Furthermore, we will consider multiscale data representation to allow coarse-to-fine data viewing and exploration. For the 2D/3D registration, DRR images can be generated using GPU.

3. The mobile C-arm positioning system has been validated using a spinal phantom. More tests on the cadaver and real patient in the clinical environment are suggested. The complex structure of human body will verify the performance of current system as well as the limitations further. We can collaborate with the hospital for the future experiments. The experienced surgeons will be recruited to evaluate the positioning system in different clinical setting, and some novice

participants are used to compare the learning curve with the case without the positioning navigation.

# Bibliography

Adelson, E. H., Simoncelli, E. P., and Freeman, W. T. (1990). Pyramids and multiscale representations. In *Proc European Conf on Visual Perception*, Paris. 47

Agarwal, S., Awan, A., and Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(11):1475–1490. 9

Ahrens, J., Law, C., Schroeder, W., Martin, K., Inc, K., and Papka, M. (2000). A parallel approach for efficiently visualizing extremely large, time-varying datasets. Technical report, Los Alamos National Laboratory. 44

Akansu, A., Haddad, R., and Caglar, H. (1993). The binomial qmf-wavelet transform for multiresolution signal decomposition. *Signal Processing, IEEE Transactions on*, 41(1):13–19. 47

Albani, J. M. and Lee, D. I. (2007). Virtual reality-assisted robotic surgery simulation. *Journal of endourology*, 21(3):285–287. 4

Ayad, M. S., Lee, J., Deguet, A., Burdette, E. C., and Prince, J. L. (2010). C-arm pose estimation using a set of coplanar ellipses in correspondence. In *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*, pages 1401–1404. IEEE. 60

Balter, J. M. and Kessler, M. L. (2007). Imaging and alignment for image-guided radiation therapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 25(8):931–7. 41

Barth, N. (1999). The gramian and k-volume in n-space: some classical results in linear algebra. *J Young Investig*, 2. 71

Belkin, M. and Niyogi, P. (2001). Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS*, volume 14, pages 585–591. 9, 75

Beyer, J., Hadwiger, M., Wolfsberger, S., and Buhler, K. (2007). High-quality multimodal volume rendering for preoperative planning of neurosurgical interventions. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1696–1703. 65

Bindal, R. K., Glaze, S., Ognoskie, M., Tunner, V., Malone, R., and Ghosh, S. (2008). Surgeon and patient radiation exposure in minimally invasive transforaminal lumbar interbody fusion: Clinical article. *Journal of Neurosurgery: Spine*, 9(6):570–573. 3

Binder, N., Bodensteiner, C., Matthäus, L., Burgkart, R., and Schweikard, A. (2006). Image guided positioning for an interactive c-arm fluoroscope. In *Computer Assisted Radiology and Surgery (CARS)*. Citeseer. 3, 7, 36

Bott, O., Dresing, K., Wagner, M., Raab, B., and Teistler, M. (2011). Informatics in radiology: use of a c-arm fluoroscopy simulator to support training in intraoperative radiography. *Radiographics*, 31(3):E65–76. 15

Bottou, L. and Bousquet, O. (2008). The tradeoffs of large scale learning. In Platt, J., Koller, D., Singer, Y., and Roweis, S., editors, *Advances in Neural Information Processing Systems*, volume 20, pages 161–168. NIPS Foundation (http://books.nips.cc). 67

Caban, J. J. and Rheingans, P. (2008). Texture-based transfer functions for direct volume rendering. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1364–1371. 87

Chaganti, S., Kumar, D., Patil, S., and Alderman, P. (2009). A language for effective communication between surgeons and radiographers in trauma theatre. *Annals of The Royal College of Surgeons of England*, 91(6):509–512. 17

Chiaverini, S., Sciavicco, L., and Siciliano, B. (1991). Control of robotic systems through singularities. In Canudas de Wit, C., editor, *Advanced Robot Control*,

volume 162 of *Lecture Notes in Control and Information Sciences*, pages 285–295. Springer Berlin / Heidelberg. 30

Chintalapani, G., Jain, A., Burkhardt, D., Prince, J., and Fichtinger, G. (2008). Ctrec: C-arm tracking and reconstruction using elliptic curves. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, pages 1–7. 60

Cleary, K. and Peters, T. M. (2010). Image-guided interventions: technology review and clinical applications. *Annual review of biomedical engineering*, 12:119–42. 1, 15, 41, 60, 65

Correa, C. and Ma, K.-L. (2008). Size-based transfer functions: A new volume exploration technique. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1380–1387. 87

Craig, J. J. (1989). *Introduction to Robotics: Mechanics and Control*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2nd edition. 26

De Castro, E. and Morandi, C. (1987). Registration of translated and rotated images using finite fourier transforms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-9(5):700–703. 7

de Moura Pinto, F. and Freitas, C. M. (2007). Design of multi-dimensional transfer functions using dimensional reduction. In *Proceedings of the 9th Joint Eurographics/IEEE VGTC conference on Visualization*, pages 131–138. Eurographics Association. 65, 87

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32:407–499. 77

Elad, M. (2010). *Sparse and redundant representations: from theory to applications in signal and image processing*. Springer. 9

Furukawa, Y., Curless, B., Seitz, S., and Szeliski, R. (2010). Towards internet-scale multi-view stereo. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1434–1441. 19

Furukawa, Y. and Ponce, J. (2010). Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376. 20

Georgii, J., Eder, M., Kovacs, L., Schneider, A., Dobritz, M., and Westermann, R. (2007). Advanced volume rendering for surgical training environments. *International Journal of Computer Assisted Radiology and Surgery*, 2:S285. 65

Gerber, N., Gavaghan, K., Bell, B., Williamson, T., Weisstanner, C., Caversaccio, M., and Weber, S. (2013). High-accuracy patient-to-image registration for the facilitation of image-guided robotic microsurgery on the head. *Biomedical Engineering, IEEE Transactions on*, 60(4):960–968. 41

Grzeda, V. and Fichtinger, G. (2010). Rotational encoding of c-arm fluoroscope with tilt sensing accelerometer. In *Proceedings of the 13th international conference on Medical image computing and computer-assisted intervention: Part III*, MICCAI'10, pages 424–431, Berlin, Heidelberg. Springer-Verlag. 37

Halíř, R. and Flusser, J. (1998). Numerically stable direct least squares fitting of ellipses. In *Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization. WSCG '98*, pages 125–132, Plzeň, Czech Republic. CZ. 26

Hladuvka, J., König, A., and Gröller, E. (2000). Curvature-based transfer functions for direct volume rendering. In *Spring Conference on Computer Graphics*, volume 16, pages 58–65. 86

Hoffmann, H. (2007). Kernel pca for novelty detection. *Pattern Recognition*, 40(3):863–874. 9

Huang, S., Cai, C., and Zhang, Y. (2011). Dimensionality reduction by using sparse reconstruction embedding. In *Advances in Multimedia Information Processing-PCM 2010*, pages 167–178. Springer. 88

Hurvitz, A. and Joskowicz, L. (2008). Registration of a ct-like atlas to fluoroscopic x-ray images using intensity correspondences. *Int. J. Computer Assisted Radiology and Surgery*, pages 493–504. 41

Hyvärinen, A. and Oja, E. (2000). Independent component analysis: algorithms and applications. *Neural networks*, 13(4):411–430. 9

Jacobs, F., Sundermann, E., De Sutter, B., Christiaens, M., and Lemahieu, I. (1998). A fast algorithm to calculate the exact radiological path through a pixel or voxel space. *CIT. JOURNAL OF COMPUTING AND INFORMATION TECHNOLOGY*, 6(1):89–94. 25

Jain, A. and Fichtinger, G. (2006). C-arm tracking and reconstruction without an external tracker. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 494–502. Springer. 60

Jain, A. K., Mustafa, T., Zhou, Y., Burdette, C., Chirikjian, G. S., and Fichtinger, G. (2005). Ftrac-a robust fluoroscope tracking fiducial. *Medical physics*, 32:3185. 60

Jerbi, T., Burdin, V., Leboucher, J., Stindel, E., and Roux, C. (2013). 2-d-3-d frequency registration using a low-dose radiographic system for knee motion estimation. *Biomedical Engineering, IEEE Transactions on*, 60(3):813–820. 7, 41

Jimenez, L. O. and Landgrebe, D. (1996). High dimensional feature reduction via projection pursuit. *ECE Technical Reports*, page 103. 9

Jolliffe, I. (2005). *Principal component analysis*. Wiley Online Library. 9

Khalifa, F., Beache, G. M., Gimel, G., Suri, J. S., and El-baz, A. S. (2011). *State-of-the-Art Medical Image Segmentation and Registration Methodologies*, volume 1. Springer US, Boston, MA. 7

Kim, H. S., Schulze, J. P., Cone, A. C., Sosinsky, G. E., and Martone, M. E. (2010). Dimensionality reduction on multi-dimensional transfer functions for multi-channel volume data sets. *Information visualization*, 9(3):167–180. 65, 87

Kindlmann, G. (2002). Transfer functions in direct volume rendering: Design, interface, interaction. *Course notes of ACM SIGGRAPH*. 65

Kindlmann, G. and Durkin, J. W. (1998). Semi-automatic generation of transfer functions for direct volume rendering. In *Proceedings of the 1998 IEEE symposium on Volume visualization*, pages 79–86. 86

Kindlmann, G., Whitaker, R., Tasdizen, T., and Moller, T. (2003). Curvature-based transfer functions for direct volume rendering: Methods and applications. In *Visualization, 2003. VIS 2003. IEEE*, pages 513–520. 86

Klein, S., Staring, M., Murphy, K., Viergever, M., and Pluim, J. P. W. (2009). elastix: A toolbox for intensity-based medical image registration. *Medical Imaging, IEEE Transactions on*, 29(1):196–205. 57

Klein, T., Benhimane, S., Traub, J., Heining, S., Euler, E., and Navab, N. (2007). Interactive guidance system for c-arm repositioning without radiation. In Horsch, A., Deserno, T., Handels, H., Meinzer, H.-P., and Tolxdorff, T., editors, *Bildverarbeitung fr die Medizin 2007*, Informatik aktuell, pages 21–25. Springer Berlin Heidelberg. 35

Kniss, J., Kindlmann, G., and Hansen, C. (2002). Multidimensional transfer functions for interactive volume rendering. *Visualization and Computer Graphics, IEEE Transactions on*, 8(3):270–285. 86

Ko, R. and Razvi, H. (2007). C-arm laser positioning device to facilitate percutaneous renal access. *Urology*, 70(2):360 – 361. 37

Kubias, A., Deinzer, F., Feldmann, T., and Paulus, D. (2007). Extended global optimization strategy for rigid 2d/3d image registration. In *Computer Analysis of Images and Patterns*, pages 759–767. Springer. 61

Ladikos, A., Benhimane, S., and Navab, N. (2008). Real-time 3d reconstruction for collision avoidance in interventional environments. In *Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention, Part II*, MICCAI '08, pages 526–534, Berlin, Heidelberg. Springer-Verlag. 36

Lang, P., Seslija, P., Chu, M., Bainbridge, D., Guiraudon, G., Jones, D., and Peters, T. (2012). Us-fluoroscopy registration for transcatheter aortic valve implantation. *Biomedical Engineering, IEEE Transactions on*, 59(5):1444–1453. 41

Lee, H., Battle, A., Raina, R., and Ng, A. Y. (2007). Efficient sparse coding algorithms. *Advances in neural information processing systems*, 19:801. 75, 77

Leschka, S., Babic, D., El Shikh, S., Wossmann, C., Schumacher, M., and Taschner, C. (2012). C-arm cone beam computed tomography needle path overlay for image-guided procedures of the spine and pelvis. *Neuroradiology*, 54:215–223. 15

Liegeois, A. (1977). Automatic supervisory control of the configuration and behavior of multibody mechanisms. *IEEE Trans. Systems, Man, and Cybernetics*, 7(12):842–868. 30

Livyatan, H., Yaniv, Z., and Joskowicz, L. (2003). Gradient-based 2-d/3-d rigid registration of fluoroscopic x-ray to ct. *Medical Imaging, IEEE Transactions on*, 22(11):1395–1406. 60

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2009). Online dictionary learning for sparse coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 689–696. ACM. 67

Markelj, P., Tomazevic, D., Likar, B., and Pernus, F. (2010). A review of 3d/2d registration methods for image-guided interventions. *Medical Image Analysis*, 1(3):192–197. 41

Matthaus, L., Binder, N., Bodensteiner, C., and Schweikard, A. (2007). Closed-form inverse kinematic solution for fluoroscopic c-arms. *Advanced Robotics*, pages 869–886. 7, 36

Maybody, M., Stevenson, C., and Solomon, S. B. (2013). Overview of navigation systems in image-guided interventions. *Techniques in Vascular and Interventional Radiology*, 16(3):136–143. 1

Meyer, C. (2000). *Matrix analysis and applied linear algebra book and solutions manual*, volume 2. Siam. 70, 71

Meyer, S. and Wolf, P. (1999). Registration of three-dimensional cardiac catheter models to single-plane fluoroscopic images. *Biomedical Engineering, IEEE Transactions on*, 46(12):1471–1479. 41

Mueller, D., Maeder, A., and O'Shea, P. (2006). The Generalised image fusion toolkit (GIFT). *Insight Journal*, pages 1–16. 47

Munbodh, R., Chen, Z., Jaffray, D. A., Moseley, D. J., Knisely, J. P., and Duncan, J. S. (2007). A frequency-based approach to locate common structure for 2d-3d intensity-based registration of setup images in prostate radiotherapy. *Medical physics*, 34:3005. 41, 61

Navab, N., Wiesner, S., Benhimane, S., Euler, E., and Heining, S. M. (2006). Visual servoing for intraoperative positioning and repositioning of mobile c-arms. In *MICCAI (1)'06*, pages 551–560. 35, 36

Nguyen, H. V., Patel, V. M., Nasrabadi, N. M., and Chellappa, R. (2012). Sparse embedding: a framework for sparsity promoting dimensionality reduction. In *Computer Vision–ECCV 2012*, pages 414–427. Springer. 9, 88

Otake, Y., Armand, M., Armiger, R., Kutzer, M., Basafa, E., Kazanzides, P., and Taylor, R. (2012). Intraoperative image-based multiview 2d/3d registration for image-guided orthopaedic surgery: Incorporation of fiducial-based c-arm tracking and gpu-acceleration. *Medical Imaging, IEEE Transactions on*, 31(4):948 –962. 60, 62

Pawiro, S., Markelj, P., Pernuš, F., and Gendrin, C. (2011). Validation for 2D/3D registration I: A new gold standard data set. *Medical Phys*, 38(3):1481–1490. xiii, 48, 52, 78

Perkins, S. and Theiler, J. (2003). Online feature selection using grafting. In *In International Conference on Machine Learning*, pages 592–599. ACM Press. 77

Pfister, H., Lorensen, B., Bajaj, C., Kindlmann, G., Schroeder, W., Avila, L. S., Raghu, K., Machiraju, R., and Lee, J. (2001). The transfer function bake-off. *Computer Graphics and Applications, IEEE*, 21(3):16–22. 65, 85

Praßni, J.-S., Ropinski, T., Mensmann, J., and Hinrichs, K. (2010). Shape-based transfer functions for volume visualization. In *Pacific Visualization Symposium (PacificVis), 2010 IEEE*, pages 9–16. IEEE. 87

Reaungamornrat, S., Otake, Y., Uneri, A., Schafer, S., Mirota, D., Nithiananthan, S., Stayman, J. W., Kleinszig, G., Khanna, A. J., Taylor, R. H., and Siewerdsen, J. H. (2012). An on-board surgical tracking and video augmentation system for c-arm image guidance. *Int. J. Computer Assisted Radiology and Surgery*, 7(5):647–665. 45, 60

Reda, F., Noble, J., Labadie, R., and Dawant, B. (2012). Automatic pre- to intra-operative ct registration for image-guided cochlear implant surgery. *Biomedical Engineering, IEEE Transactions on*, 59(11):3070–3077. 41

Reddy, B. and Chatterji, B. N. (1996). An fft-based technique for translation, rotation, and scale-invariant image registration. *Image Processing, IEEE Transactions on*, 5(8):1266–1271. 7

Rohlfing, T., Russakoff, D. B., Denzler, J., Mori, K., and Maurer Jr, C. R. (2005). Progressive attenuation fields: fast 2d-3d image registration without precomputation. *Medical physics*, 32:2870. 41, 61

Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326. 9

Ruijters, D., ter Haar-Romeny, B. M., and Suetens, P. (2008). Efficient gpu-accelerated elastic image registration. In *Proceedings of the Sixth IASTED International Conference on Biomedical Engineering*, BioMED '08, pages 419–424, Anaheim, CA, USA. ACTA Press. 62

Sadowsky, O., Chintalapani, G., and Taylor, R. H. (2007). Deformable 2d-3d registration of the pelvis with a limited field of view, using shape statistics. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007*, pages 519–526. Springer. 60

Salama, C. R., Keller, M., and Kohlmann, P. (2006). High-level user interfaces for transfer function design with semantics. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):1021–1028. 87

Schaller, C., Rohkohl, C., Penne, J., Stürmer, M., and Hornegger, J. (2009). Inverse c-arm positioning for interventional procedures using real-time body part detection. In *Proceedings of the 12th International Conference on Medical Image Computing*

*and Computer-Assisted Intervention: Part I*, MICCAI '09, pages 549–556, Berlin, Heidelberg. Springer-Verlag. 36

Seitz, S. M., Curless, B., Diebel, J., Scharstein, D., and Szeliski, R. (2006). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 519–528. IEEE. 19

Sereda, P., Bartroli, A., Serlie, I. W. O., and Gerritsen, F. (2006). Visualization of boundaries in volumetric data sets using lh histograms. *Visualization and Computer Graphics, IEEE Transactions on*, 12(2):208–218. 86

Shapiro, A. and Wardi, Y. (1996). Convergence analysis of gradient descent stochastic algorithms. *Journal of Optimization Theory and Applications*, 91(2):439–454. 81

Siciliano, B. (1990). A closed-loop inverse kinematic scheme for on-line joint-based robot control. *Robotica*, 8(03):231–243. 30

Snavely, N., Seitz, S. M., and Szeliski, R. (2006). Photo tourism: exploring photo collections in 3d. *ACM Trans. Graph.*, 25(3):835–846. 19

Stabile, G., Scaglione, M., del Greco, M., De Ponti, R., Bongiorni, M. G., Zoppo, F., Soldati, E., Marazzi, R., Marini, M., Gaita, F., et al. (2012). Reduced fluoroscopy exposure during ablation of atrial fibrillation using a novel electroanatomical navigation system: a multicentre experience. *Europace*, 14(1):60–65. 3

Starck, J.-L., Elad, M., and Donoho, D. L. (2005). Image decomposition via the combination of sparse representations and a variational approach. *Image Processing, IEEE Transactions on*, 14(10):1570–1582. 9

Su, P., Yang, J., Lu, K., Yu, N., Wong, S., and Xue, Z. (2013). A fast ct and ct-fluoroscopy registration algorithm with respiratory motion compensation for image-guided lung intervention. *Biomedical Engineering, IEEE Transactions on*, 60(7):2034–2041. 41

Synowitz, M. and Kiwit, J. (2006). Surgeon's radiation exposure during percutaneous vertebroplasty. *Journal of Neurosurgery: Spine*, 4(2):106–109. 3

Takanashi, I., Lum, E. B., Ma, K.-L., and Muraki, S. (2002). Ispace: Interactive volume data classification techniques using independent component analysis. In *Computer Graphics and Applications, 2002. Proceedings. 10th Pacific Conference on*, pages 366–374. IEEE. 87

Tenenbaum, J. B., De Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323. 9

Tietjen, C., Isenberg, T., and Preim, B. (2005). Combining silhouettes, surface, and volume rendering for surgery education and planning. In *Proceedings of the Seventh Joint Eurographics/IEEE VGTC conference on Visualization*, pages 303–310. Eurographics Association. 65

van de Kraats G.P. Penney D. Tomazevic Th. van Walsum, E. and Niessen, W. (2005). Standardized evaluation methodology for 2D-3D registration. *Medical Imaging, IEEE Transactions on*, 24(9):1177–1190. 41, 53

van der Bom, M., Bartels, L., Gounis, M., Homan, R., Timmer, J., Viergever, M., and Pluim, J. (2010). Robust initialization of 2D-3D image registration using the projection-slice theorem and phase correlation. *Medical Physics*, 37(4):1884–1892. 8, 52

Van der Maaten, L., Postma, E., and Van Den Herik, H. (2009). Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10:1–41. 8, 65, 87

van der Merwe, B. (2012). Radiation dose to surgeons in theatre. *South African journal of surgery. Suid-Afrikaanse tydskrif vir chirurgie*, 50(2):26–9. 21

Wang, J. (2011). Classical multidimensional scaling. In *Geometric Structure of High-Dimensional Data and Dimensionality Reduction*, pages 115–129. Springer. 9

Wang, J., Fallavollita, P., Wang, L., Kreiser, M., and Navab, N. (2012a). Augmented reality during angiography: integration of a virtual mirror for improved 2d/3d visualization. In *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*, pages 257–264. IEEE. 65

Wang, L., Fallavollita, P., Zou, R., Chen, X., Weidert, S., and Navab, N. (2012b). Closed-form inverse kinematics for interventional c-arm x-ray imaging with six degrees of freedom: Modeling and application. *Medical Imaging, IEEE Transactions on*, 31(5):1086 –1099. 7, 15, 35

Wright, J., Ma, Y., Mairal, J., Sapiro, G., Huang, T. S., and Yan, S. (2010). Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, 98(6):1031–1044. 88

Xiaogang, D., Jianwu, D., Yangping, W., Xinguo, L., and Sha, L. (2013). An algorithm multi-resolution medical image registration based on firefly algorithm and powell. In *Intelligent System Design and Engineering Applications (ISDEA), 2013 Third International Conference on*, pages 274–277. IEEE. 41, 61

Xu, R. and Chen, Y. (2007). Wavelet-based multiresolution medical image registration strategy combining mutual information with spatial information. *International journal of innovative computing, information and control*, 3(2):285–296. 47, 61

Yoo, T. S., Ackerman, M. J., and Lorensen, W. E. (2002). Engineering and algorithm design for an image processing API: A technical report on itk-the insight toolkit. *Proc. of Medicine Meets Virtual Reality*, pages 586–592. 47

Zhang, L., Yang, M., Feng, Z., and Zhang, D. (2010). On the dimensionality reduction for sparse representation based face recognition. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 1237–1240. IEEE. 9, 88

Zhao, X. and Kaufman, A. (2010). Multi-dimensional reduction and transfer function design using parallel coordinates. In *Proceedings of the 8th IEEE/EG international conference on Volume Graphics*, pages 69–76. Eurographics Association. 87

# Vita

Zhenzhou Shao was born in Laiwu, Shandong, China in 1985. After graduation from Northeastern University in Shenyang, China with the Bachelor of Electrical Engineering degree, he headed for Michigan Technological University in Houghton, MI, USA in 2009 for further graduate study in the PhD program. He accepted a graduate teaching assistantship from the Department of Electrical and Computer Engineering from 2009 to 2012, and during those years, he became a PhD candidate. He moved to Knoxville, TN in February 2012 and enrolled in the University of Tennessee at Knoxville in Fall semester of 2012. He graduated with a Doctor of Philosophy degree in Biomedical Engineering in December 2013.