

**PROJECTION-BASED SPATIAL
AUGMENTED REALITY FOR
INTERACTIVE VISUAL GUIDANCE IN
SURGERY**

WEN RONG

NATIONAL UNIVERSITY OF SINGAPORE

2013

**PROJECTION-BASED SPATIAL
AUGMENTED REALITY FOR
INTERACTIVE VISUAL GUIDANCE IN
SURGERY**

WEN RONG

(B.Eng., M.Sc., Chongqing University, Chongqing, China)

A THESIS SUBMITTED FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MECHANICAL ENGINEERING
NATIONAL UNIVERSITY OF SINGAPORE

2013

Declaration

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

A handwritten signature in black ink, appearing to read 'Wen Rong', with a stylized, cursive script.

Wen Rong

10 January 2013

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisors, Dr. CHUI Chee Kong and Assoc. Prof. LIM Kah Bin, for your constant guidance, motivation and untiring help during my Ph.D. candidature. Without your insights and comments, this thesis and other publications of mine would not have been possible. Thanks for your kind understanding, support and encouragement during my life in Singapore. For everything you have done for me, I can say that I am very lucky to be your student and to work with you.

I would like to sincerely thank the members in the panel of my Oral Qualifying Examination (QE), Assoc. Prof. TEO Chee Leong from the Department of Mechanical Engineering (ME, NUS) and Assoc. Prof. ONG Sim Heng from the Department of Electrical & Computer Engineering (ECE, NUS). Thanks for your sound advices and good ideas proposed in the QE examination. My thanks also go to Dr. CHANG Kin-Yong Stephen from the Department of Surgery, National University Hospital (NUH), who gave me great help in the animal experiments with a senior surgeon's point of view. Without their guidance and mentorship, it would not have been possible for me to accomplish such an interdisciplinary work.

I had a good time with my group members. It is my pleasure to acknowledge all my current and previous colleagues including Mr. YANG Liangjing, Mr. HUANG

Wei Hsuan, Mr. CHNG Chin Boon, Dr. QIN Jing from the Chinese University of Hong Kong (CSE, CUHK), Dr. NGUYEN Phu Binh (ECE, NUS), Mr. LEE Chun Siong, Mr. WU Jichuan, Mr. XIONG Linfei, Ms. HO Yick Wai Yvonne Audrey, Mr. DUAN Bin, Mr. WANG Gang, Ms. WU Zimei and many others. Thanks for your generous help and invaluable advices. Most importantly, your friendship made all these unforgettable experiences for me.

I would like to thank Dr. LIU Jiang Jimmy, Dr. ZHANG Jing and Mr. YANG Tao, from the Institute for Infocomm Research(I2R), Agency for Science, Technology and Research (A*STAR). I will always be grateful to your kind supports during my tough times.

It is really my honour to work in the Control & Mechatronics Laboratory. My sincere thanks go to the hard-working staff in this laboratory, Ms. OOI-TOH Chew Hoey, Ms. Hamidah Bte JASMAN, Ms. TSHIN Oi Meng, Mr. Sakthiyavan KUPPUSAMY and Mr. YEE Choon Seng. All of them are being considerate and supportive.

My thanks go to the Department of Mechanical Engineering, who offered me the generous scholarship and enabled me to concentrate on the thesis researches during the candidature. Many special thanks are extended to the staff working in the department office, Ms. TEO Lay Tin Sharen, Ms. Helen ANG and many others.

Last but not least, I would like to thank all of my family members for their love, encouragement and sacrifice. I am deeply thankful to my parents who raised me and supported me in all my pursuits, to my parents-in-law who took charge of many family matters when I and my wife were away from home. My special thanks go to my love, Ms. FU Shanshan who always expresses her endless support,

inspiration and faith in me. Without their consideration and endless supports, I would not be able to devote myself to this doctoral programme.

Wen Rong

1 January, 2013

Contents

| | |
|---|--------------|
| Summary | IX |
| List of Figures | XII |
| List of Tables | XVIII |
| List of Abbreviations | XIX |
| 1 INTRODUCTION | 1 |
| 1.1 From Virtual Reality to Augmented Reality | 1 |
| 1.2 Medical Augmented Reality | 6 |
| 1.3 Research Objectives and Contributions | 8 |
| 1.4 Thesis Organization | 10 |
| 2 LITERATURE REVIEW | 12 |
| 2.1 ProCam System | 12 |

| | | |
|----------|--|-----------|
| 2.2 | ProCam Calibration | 14 |
| 2.2.1 | Camera and Projector Calibration | 14 |
| 2.2.2 | System Calibration | 17 |
| 2.3 | Projection Correction | 20 |
| 2.4 | Registration in Augmented Reality Surgery | 23 |
| 2.4.1 | AR Registration | 25 |
| 2.4.2 | Registration in Image-guided Surgery | 27 |
| 2.5 | Human-computer Interaction in VR and AR Environment | 30 |
| 2.5.1 | HCI Design and Methods | 31 |
| 2.5.2 | Augmented Interaction | 33 |
| 2.6 | Summary | 35 |
| 3 | SYSTEM CALIBRATION | 37 |
| 3.1 | Camera and Projector Calibration | 37 |
| 3.2 | Calibration for Static Surface | 39 |
| 3.3 | Calibration for Dynamic Surface | 41 |
| 3.3.1 | Feature Initialization in Camera Image | 44 |
| 3.3.2 | Tracking of Multiple Feature Points with Extended Kalman Filter | 47 |

| | | |
|----------|--|-----------|
| 3.3.3 | Feature Point Matching Based on Minimal Bending Energy | 53 |
| 3.4 | Summary | 56 |
| 4 | GEOMETRIC AND RADIOMETRIC CORRECTION | 57 |
| 4.1 | Geometric Correction | 58 |
| 4.1.1 | Principle of Viewer-dependent Pre-warping | 59 |
| 4.1.2 | Piecewise Pre-warping | 61 |
| 4.2 | Radiometric Correction | 65 |
| 4.2.1 | Radiometric Model for ProCam | 65 |
| 4.2.2 | Radiometric Compensation | 68 |
| 4.3 | Texture Mapping for Pixel Value Correction | 71 |
| 4.4 | Summary | 73 |
| 5 | REGISTRATION | 74 |
| 5.1 | Registration between Surgical Model and Patient Body | 75 |
| 5.1.1 | Data Acquisition and Preprocessing | 75 |
| 5.1.2 | Surface Matching for Optimal Data Alignment | 79 |
| 5.2 | Registration between Model-Projection Image and Patient Body | 82 |
| 5.3 | Summary | 85 |

| | | |
|----------|---|------------|
| 6 | AUGMENTED INTERACTION | 87 |
| 6.1 | Preoperative Planning | 88 |
| 6.2 | Interactive Supervisory Guidance | 93 |
| 6.3 | Augmented Needle Insertion | 97 |
| 6.4 | Summary | 101 |
| 7 | EXPERIMENTS AND DISCUSSION | 103 |
| 7.1 | Projection Accuracy Evaluation | 105 |
| 7.2 | Registration Evaluation | 108 |
| 7.3 | Evaluation of Augmented Interaction | 110 |
| 7.4 | Parallel Acceleration with GPU | 115 |
| 7.5 | Summary | 118 |
| 8 | CONCLUSION | 120 |
| 8.1 | Summary of Contributions | 121 |
| 8.2 | Future Work | 123 |
| | Bibliography | 125 |
| | List of Publications | 142 |

Summary

Computer-assisted surgery (CAS) has tremendously challenged traditional surgical procedures and methods with its advantages of modern medical imaging, presurgical planning using accurate three-dimensional (3D) surgical models and computer controlled robotic surgery. Medical images including preoperative or intraoperative images are employed as guidance to assist surgeons to track the surgical instrument and target anatomy structures during surgery. Most image-guided surgical treatments are minimally invasive. However, the image guidance procedure is constrained by the indirect image-organ registration and limited visual feedback of interventional results. Augmented Reality (AR) is an emerging technique enhancing display integration of computer-generated images and actual objects. It can be used to extend surgeons' visual perception of the anatomy and surgical tools that are beneath the real surgical scene.

This work introduces a projector-camera (ProCam) system into the surgical procedure to develop a direct AR based surgical planning and navigation mechanism. Through overlaying the projection of planning data on the specific position of the patient (skin) surface, surgeons can directly supervise robot-assisted execution according to the presurgical planning. New solutions are proposed to overcome the existing visual and operational limitations in the image-guided surgery (IGS), and specifically, in the IGS which integrates a robotic assistance. Clinical

viability of this advanced human-computer interaction (HCI) is investigated via *ex vivo* and *in vivo* experiments.

Calibration methods for the ProCam system were investigated to establish an accurate pixel correspondence between the projector and camera image. A phase shifted structured pattern was used for pixel encoding and decoding. Projection on an arbitrary surface was subjected to geometric and radiometric distortion. In order to minimize geometry distortion caused by surface variance, *an improved piecewise region based texture mapping correction method* was proposed. Since radiometry distortion was mainly due to angle of projection, surface texture and lightings, *radiometric model based image compensation* was developed to restore the projected image from form factor, screen color and environmental lighting.

Registration is a challenging problem especially when the projector-based AR is used for navigation in a surgical environment. Projection with accurate models' profile on the specific region of the patient surface is essential in surgery. *A new registration method using surface matching and point-based registration algorithms* was developed for patient-model and patient-world registration respectively.

A further study was conducted on a direct augmented interaction with dynamic projection guidance for surgical navigation. With stereoscopic tracking and fiducial marker based registration, surgical intervention within the patient body was displayed through the real surgical tool interacting with the overlaying computer-generated models. *Viewer-dependent model-world registration* enabled the virtual surgical tool model to match the corresponding real one accurately in the world space. *A 3D structured model based hand gesture recognition* was developed for surgeon-AR interaction. This innovative hand-gesture control provides the surgeon an efficient means to directly interact with the surgical AR environment without contact infection. In addition, this study explores *projection-based*

visualization for robot-assisted needle insertion. Operation of the surgical robot was integrated into the AR environment.

Interactive visual guidance with projector-based AR enables computer-generated surgical models to be directly visualized and manipulated on the patient's skin. It has advantages of consistent viewing focus on the patient, extended field of view and improved augmented interaction. The proposed AR guidance mechanism was tested in surgical experiments with percutaneous robot-assisted radiofrequency (RF) needle insertion and direct augmented interaction. The experimental results on the phantom and porcine models demonstrated its clinical viability in the robot-assisted RF surgery.

List of Figures

| | | |
|-----|---|----|
| 1-1 | Modality of working environment: traditional (a), VR (b) and AR (c) environment. | 2 |
| 1-2 | VR environment (a) Sensorama (Kock, 2008) (b) Ford's Cave Automated Virtual Environment (CAVE) is used to evaluate the prototype design of a new car (Burns, 2010). | 3 |
| 2-1 | ProCam system (a) High-speed ProCam system for 3D measurement (Toma, 2010). (b) ProCam system with LCD projector and digital camera for keystone correction on the presentation screen (Sukthankar et al., 2000). | 13 |
| 2-2 | Geometric relationship between the 3D object point P_w in the world coordinate system and its corresponding point P_d in the image coordinate system (Salvi et al., 2002). | 15 |
| 2-3 | Sequential binary-coded pattern projection (a) and color stripe indexing (b) used to establish pixel correspondence in ProCam calibration (Geng, 2011). | 19 |

| | | |
|-----|---|----|
| 2-4 | Registration with clip-on marker set and reference frame for MR imager based AR surgical guidance in a biopsy experiment. (Wacker et al., 2006). | 24 |
| 2-5 | Neural network used for classification of gestures features (Kulkarni and Lokhande, 2010). | 32 |
| 2-6 | Augmented interaction with attached devices and equipments (Nee et al., 2012). | 34 |
| 3-1 | Encoding of projector images and decoding of camera images. | 39 |
| 3-2 | Projection of binary coded patterns with phase shift variation. (a) A binary-coded pattern of two strips. (b) A binary-coded pattern of twenty-four strips with phase shift. | 41 |
| 3-3 | Workflow of the hybrid algorithm. | 43 |
| 3-4 | Projection image on the patient surface (a) and its corresponding edge map with surface feature points (b). | 46 |
| 3-5 | New mapping establishment for the ProCam system. \mathbf{M}_{p-co} is the initial pixel mapping between the projector and camera image. P_{co} are the initial corresponding points of the projector image points P_p on the camera image. The surface feature points P_s (green points) have their corresponding points P_i (yellow points) on the camera image. The new corresponding points of P_i can be found by 2D lookup table from the mapping \mathbf{M}_{p-co} . For the P_i without corresponding points on the camera image, nearest neighbour interpolation is used to find their correspondences P_{co} . | 46 |

| | | |
|-----|--|----|
| 3-6 | Patterns for motion vectors grouping. Motion vectors grouping (b) depends on motion motivated by the internal force under the surface (a). | 49 |
| 3-7 | Motion field prediction with uncertainty error. The red points represent the motion centers of the different feature groups. The green regions represent the prediction regions with uncertainty error. . . . | 51 |
| 4-1 | Geometric and radiometric distortion. | 58 |
| 4-2 | Viewer-dependent geometric correction. | 61 |
| 4-3 | Geometric correction on a planar surface (b) and curved surface (c) with piecewise pre-warping method. The piecewise regions are defined by the four feature points in quadrilaterals (a) in this example. . . . | 64 |
| 4-4 | Radiometric correction of a liver model on a mannequin body: (a) before correction, (b) after correction. | 70 |
| 4-5 | Blob cluster are projected to establish the texture mapping. | 71 |
| 4-6 | Projection correction on a curved surface based on texture mapping: (a) projection distortion (checkerboard pattern) caused by the curved surface; (b) projection on the curved surface. | 73 |
| 5-1 | Data acquisition for registration. | 75 |
| 5-2 | Geometry for retrieving the surface data with a projector-camera system. | 77 |

| | | |
|-----|--|----|
| 5-3 | Surface matching based registration between patient's surface point cloud (a) and surface model (b). The blue regions represent the matched surface data. | 81 |
| 5-4 | Marker-based registration for SAR (M_1, M_2, M_3 are three markers attached onto the mannequin body.) | 82 |
| 5-5 | Geometric correction and registration of the model-projection on an irregular surface with its corresponding internal object. | 84 |
| 6-1 | Work flow of the proposed interface for an AR-guided surgery. . . . | 88 |
| 6-2 | Construction of the optimal ablation model (brown figure: virtual construction of tumor; green dots: designated location of needle tips; red wireframe: predicted ablation region; blue: resultant necrosis region). | 92 |
| 6-3 | Surgical model (ablation model) based surgical planning: (a) ablation model planning is based on anatomic models; (b) path planning is based on available workspace of the surgical robot. | 93 |
| 6-4 | 3D model based hand gesture recognition. (a) 3D graphic model of hand gesture and its corresponding 2D processed image. (b) Hand plane derived from the spatial point cloud of a hand gesture. (c) Key geometric parameters of the hand gestures | 95 |
| 6-5 | (a) ProCam-based augmented needle insertion (b) Initialization of the surgical robotic system in an operating room. | 97 |
| 6-6 | Transformation between the different workspaces for intraoperative augmented needle insertion. | 98 |

| | | |
|-----|--|-----|
| 7-1 | ProCam AR guidance system: (a) system model overview; (b) snapshot of the setup in the laboratory. | 104 |
| 7-2 | Mannequin with a removable lid and plasticine models inside. (a) With the lid in place for projection examination. (b) With the lid removed and plasticine models exposed for insertion verification. . . | 104 |
| 7-3 | Deploying markers on the porcine surface before CT scanning (a) and surgical planning based on porcine anatomy model (b). | 105 |
| 7-4 | Projection ((a) distorted (b) corrected) on the mannequin. | 106 |
| 7-5 | Projection of a checkerboard pattern on a dynamic blank paper. . . | 107 |
| 7-6 | Examination of model-patient registration by overlaying the plasticine models on the real ones which were placed inside the mannequin. Projection of the image with the real plasticine models captured from the real camera's view (left) was considered as projection with expected position. Projection of the registered virtual plasticine models captured from the virtual camera's view in the anatomic model space (right) was tested. | 108 |
| 7-7 | Registration errors of the four plasticine models. | 109 |
| 7-8 | Spatial AR based visual guidance. (a) AR display of planning data on the porcine belly. (b) Surgeons can provide their feedback based on the AR interface. | 111 |

| | | |
|------|---|-----|
| 7-9 | ProCam-based surgical AR display on the mannequin body. (a) AR display of the critical structures, vessels and tumor. (b) The preplanned insertion point and the insertion trajectory were highlighted on the patient surface for the first needle implant. | 111 |
| 7-10 | Viewer's position dependent AR display of the needle path for augmented interaction between the real and virtual needle segments. . | 112 |
| 7-11 | Augmented needle insertion process: (a)-(b) direct augmented interaction with the RF needle insertion providing surgeon's visual feedback for supervision of robotic execution; (c) insertion completed with the overlapping ablation model. | 113 |
| 7-12 | Comparison of the actual trajectory of the RF needle insertion (a) with its preplanned one generated in the preoperative planning (b). | 113 |
| 7-13 | Needle implants for the tumor model test. The red crosses represent the expected needle placements. | 114 |
| 7-14 | Parallel matrix operation based on CUDA structure for accelerating EFK tracking and bending energy minimization. | 117 |
| 7-15 | Performance comparison: (a) comparative graphic of matrix operation in the process of EFK computation and bending energy minimization with CUDA vs. CPU; (b) comparative graphic of edge-map generation with CUDA vs. CPU. | 118 |

List of Tables

| | | |
|-----|--|-----|
| 1.1 | Comparison among different AR technologies | 5 |
| 5.1 | Data spaces used for ProCam-baed AR construction | 76 |
| 7.1 | Deviation statistics for projection correction | 106 |
| 7.2 | Experimental data for robot-assisted needle insertion. | 114 |

List of Abbreviations

| | |
|--------|--|
| 3D | Three-dimensional |
| AR | Augmented Reality |
| CAS | Computer-assisted Surgery |
| CFC | Candidates of the feature correspondence |
| CT | Computed Tomography |
| CUDA | Compute Unified Device Architecture |
| EKF | Extended Kalman Filter |
| EPF | Error of the predict projection field |
| GPU | Graphics Processing Unit |
| HCI | Human-computer Interaction |
| HMD | Head-mounted Display |
| IGS | Image-guided Surgery |
| MIS | Minimally Invasive Surgery |
| MRI | Magnetic Resonance Imaging |
| ProCam | Projector-camera |
| RF | Radiofrequency |
| RFA | Radiofrequency Ablation |
| SAR | Spatial Augmented Reality |
| VR | Virtual Reality |

Chapter 1

INTRODUCTION

It is human nature to explore the world by simulation, for fun or for learning. Ever since the prehistoric ages, our primitive ancestors have started to "reconstruct" the natural creatures. The cavemen sat around the fire producing animal images cast on the cave wall with shadows made from their bodies. They played with these shadows, fabricating the earliest human legends. Today, humans have gone through thousands of years of evolution. However, that inner nature has never been changed but developed. Now, we want to create a new dream world combining reality with virtuality.

1.1 From Virtual Reality to Augmented Reality

In the common real environment, a gap is consistently existing between the actual reality and the computer-generated information (data, images and models)(Figure 1-1 (a)). The real and virtual information thus cannot be timely and spatially shared with each other. Virtual reality (VR) is a technology that creates a digital environment to simulate physical presence in the actual and imaginary worlds. It

eliminates the gap within a purely virtual environment (Figure 1-1 (b)). VR has been driven by computer simulation technology since Morton Heilig started his first invention on Sensorama in 1957 (Figure 1-2a), a simulator providing users an experience of riding a motorcycle (Kock, 2008). Based on the 3D motion pictures, Sensorama could be used to simulate driving sensation of the riders.

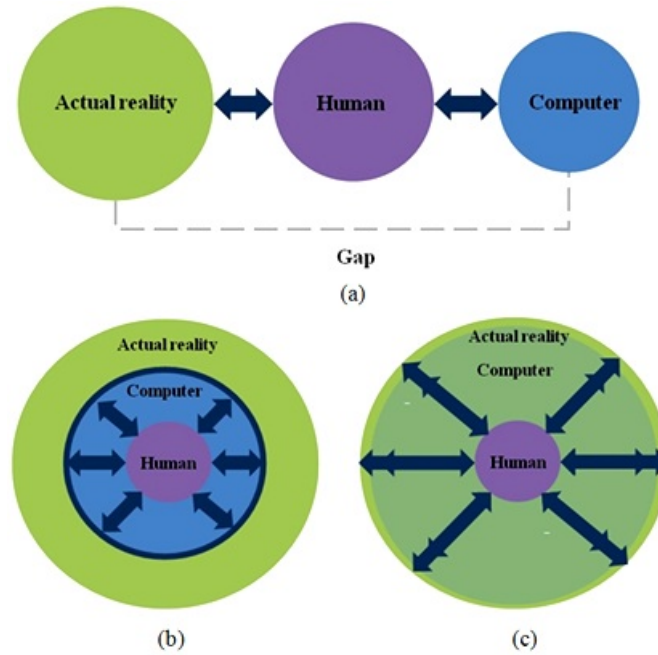


Figure 1-1. Modality of working environment: traditional (a), VR (b) and AR (c) environment.

Users in a VR environment can sense visual-dominant feedback through various sensors including display screen for visual perception and audio and haptic devices for hearing and operational sensing. In order to simulate the objects and states of real world, modelling is an important process: the regulations followed by the objects, object relationships, interactions between objects, and development and change in the real world are reflected as various data in digital space for presentation (Zhao, 2009). With development of modern multimedia technologies, current VR technology is used in a broad range of applications such as games, movies, designing and training (Figure 1-2b). The visual-dominant simulation enables people to be safely and friendly interact with the virtual objects with

image-guided information that may not exist in the actual world.

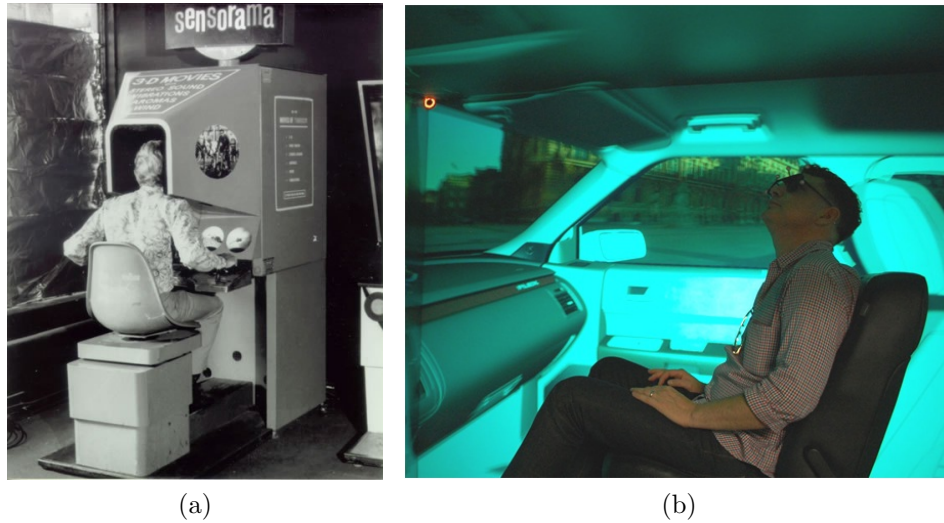


Figure 1-2. VR environment (a) Sensorama (Kock, 2008) (b) Ford's Cave Automated Virtual Environment (CAVE) is used to evaluate the prototype design of a new car (Burns, 2010).

Although VR can be used to realize a virtual environment construction, there may be a lack of information linkage between the virtual objects and their corresponding real scene. The virtual and physical world scenes are separated, and there is no data flow between them (Figure 1-1 (b)). To integrate these two paralleled environments, augmented reality (AR) was proposed to superimpose computer-generated images onto the user's view of the real scene. It enables users to simultaneously perceive additional information generated from the virtual scenes. In this way, AR eliminates the gap in Figure 1-1 by augmenting the real environment with a synthetic virtual information. The augmented information could establish a strong link to the real environment especially on spatial relation between the augmentations and the real environment. Compared to VR technology, AR is characterized as fusion of real and virtual data within the real world environment rather than solely relying on the artificially created virtual environment.

According to display modality of synthesizing virtual and real information, AR technology can be categorized as screen-based augmentation, optical see-through

based augmentation and spatial augmentation (Bimber and Raskar, 2005). Direct augmentation also called spatial augmented reality (SAR) is using projector-camera (ProCam) system or hologram imaging system to present virtual information directly in the real world environment. Based on the above different approaches of augmentations, four kinds of AR systems are mostly used: monitor-based display system, head-mounted display (HMD) device, semi-transparent mirror system and ProCam system.

Table 1.1 shows a comparison among these AR technologies. From Table 1.1, we observe that projector-based spatial AR offers the distinct advantages of better ergonomics, large field of view that allows users not wearing heavy helmet, consistent viewing focus, improved augmented interaction and flexible environmental adaptability. However, its use is challenged by directly overlaying the projection images onto the actual scenes to achieve an "actual mergence" of virtual-real information in the physical world rather than image overlaying each other on the screens or monitors.

As the real world scene is augmented by computer's synthetic information, the traditional style of human-computer interaction (HCI) has to be changed accordingly to adapt to this information-fusion environment. To eliminate the large gap between the computer and the real world in the traditional HCI, augmented interaction is brought into AR environment as a new style of HCI (Rekimoto and Nagao, 1995). Augmented interaction aims to fuse HCI between human and computer, and between human and actual world together. The computer's role is to assist and enhance interactions between humans and the real world working as transparent as possible. The user's focus will thus not be on the computer, but on the augmented real world. In this way, users can simultaneously interact with the virtual objects in a real scene, which provides users both real and virtual

Table 1.1: Comparison among different AR technologies

| | Advantages | Disadvantages |
|--|--|---|
| Desktop computer, Handheld device | Easier hardware settings; | Limited field of view; |
| | Easier virtual-real scene merging; | Indirect real-scene perception and interaction; |
| | Fast project development. | Limited mobility. |
| Semi-transparent mirror, HMD device | Better immersion than monitor-based display; | Limited field of view; |
| | Brighter and higher-resolution image; | Indirect real-scene perception and interaction; |
| | Consistent viewing focus; | Ergonomical problems (wearing HMD device or using visual assistance); |
| | Improved mobility. | Number of observers is restricted. |
| Projector-camera, Hologram | High immersion on the physical objects; direct real scene perception | Difficult system calibration; |
| | Direct augmented interaction; | Shadow-casting of the physical objects. |
| | Extendable field of view; | |
| | Good ergonomics. | |

information at the same time.

1.2 Medical Augmented Reality

Minimally Invasive Surgery (MIS) is a surgical procedure performed through small artificial incisions with specially designed surgical instruments, instead of creation of large access trauma to expose the relevant anatomy. Compared with traditional open surgery, MIS offers advantages of minimizing invasiveness, including the reduction of tissue trauma, intraoperative blood loss, risk of post operative infection, pain experienced by the patient and recovery time (Amanatullah et al., 2012). However, indirectly accessed operation in MIS may cause problems such as restricted vision and difficult hand-eye coordination. Developing AR technology with modern optics, computer graphics, computer vision and robotics provides possible chances to resolve the problems described above. Image-guided surgery (IGS) is assuming an increasingly important role, especially with the current emphasis on MIS procedures (Terence, 2001). Development of modern medical imaging technology (e.g. Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and Ultrasound) enables VR and AR to provide an interactive guidance in IGS and MIS.

Medical virtual reality can construct computer-generated models based virtual surgical environment including reconstruction of anatomical and pathological structures as well as simulation of virtual surgical operation. With assistance of various sensors attached on surgeons and surgical tools, more extensive visualization and exploitation can be then led for diagnosis support or surgical preplanning by model immersion, interaction and navigation. However, the medical virtual reality is limited to the purely surgical model simulation without exploiting the real surgical field (Figure 1-1 (b)). Therefore, it is only used for surgical training or

planning (Soler et al., 2004).

Medical augmented reality has brought new visualization and interaction solutions into perspective. The introduction of AR to surgical treatment creates a virtual medium between preoperative surgical plan and intraoperative environment. Due to the advancement in tracking, visualization and display technology, computer-aided medical procedure based AR solution was examined in the context of MIS (Navab et al., 2007). AR system has been developed to enhance the endoscopic or laparoscopic view and enables surgeons to view hidden critical structures (e.g. arteries or nerves), pathologies (e.g. tumors), risk regions or the results of a preoperative planning such as pathways, trajectories or distances (Konishia et al., 2005). These data will be shown as if they were beneath the surface of the surgical scene and hence more intuitive. With HMD based AR interface, MRI-guided tumour extraction and therapy could be more efficient (Liao et al., 2010). Clinical testing of projector-based visualization system was reported which showed its potential to develop into a surgical navigation systems (Krempien et al., 2008). However, current medical AR based surgical guidance is constrained by the following problems.

Firstly, surgeons' mobility and field of view are limited during the surgery. Surgeons might suffer from ergonomics problems such as wearing the heavy HMD devices, tracking sensors and cables. Multiple observers are not allowed for the optical see-through based AR display.

Secondly, since there is an AR display device constantly existing between surgeons and the real surgical field, indirect real-scene perception and interaction in the current medical AR guidance system may cause problems of difficulties in hand-eye coordination and difficulties in incorporating surgical tools and robot-assistance into image-based surgical guidance. The indirect and closed AR inter-

face may limit visual feedback of augmented interaction during the surgery.

Thirdly, manual registration is mostly used in the current IGS as well as medical AR guidance. Accuracy of needle insertion is limited to the range of CT slices in the single-slice registration procedure which is used in the semi-transparent AR guidance (Fichtinger et al., 2005). Tracking and timing synchronization might be a problem in HMD-based AR guidance.

Last but not least, sterilization is one of surgeon’s most concerned problems in surgery. The hand-held monitor and panel operation based AR display may not be favored by surgeons due to sterilization issues.

1.3 Research Objectives and Contributions

Based on the above problems in the existing medical AR based IGS, we aim to find new methods and algorithms that enable surgeons to supervise percutaneous surgical augmented intervention. Surgeons should be able to directly detect the augmented medical information on the patient body, and directly interact with the augmented medical data and surgical models.

The research objective in this study is to establish a surgical AR guidance mechanism that provides surgeons, via projection images of the surgical models, with direct visual feedback and interaction for intraoperative supervision of robotic RF needle insertion. It covers the following aspects.

Correct geometric and radiometric projection distortion to construct an immersive ProCam-based surgical AR environment. Although some research groups have been working on the calibration methods and projection correction for ProCam system (Salvi et al., 2002, Wang et al., 2010c), it is not an easy task to correct

projection images overlaying on an arbitrary patient belly surface with geometric and radiometric corrections, especially on a dynamic patient body due to his or her free breathing.

Register the surgical models and projection images with patient body. Registration is always a challenging problem in AR technology, which aims to accurately synthesize the virtual augmented information within the real environment. In this study, we have to overlay the surgical models that are generated in the pre-operative surgical planning onto their corresponding regions on the patient body with correct view perspectives. In addition to the surgical model registration, the needle registration is another important problem in this study, which attempts to solve coincidence between the real and virtual needle during the needle insertion.

Realize direct augmented interaction between surgeons and surgical AR environment, and augmented percutaneous needle insertion by surgical robot. Projector-based AR has the advantage of direct augmented interaction because of its large field of view and open user interface. This enables mergence of virtual and actual objects in the real world without any display device between the users and the objects. However, this immersive image mergence may cause difficulties in both registration and augmented interactions. The AR environment should be correctly reconstructed on the real object surfaces rather than displaying combined virtual and real scene images on a computer screen. The users thus would have to directly interact with both real and projection-based virtual objects simultaneously in the real world. The surgical robotic needle insertion can also be integrated into the ProCam-based AR environment due to its open user interface. In order to provide surgeons a direct user feedback in this projection-based AR environment, we have been trying to develop a new hand-gesture based method for human-computer interaction (HCI). As to augmented interaction between AR and surgical tools,

the augmented interaction is studied among the needle, projection, virtual models and patient body in process of the robotic needle insertion.

The contribution of this research work lies in *providing surgeons a new way to supervise percutaneous AR-based IGS, which overcomes the existing visual and operational limitations in the minimally invasive surgery (MIS)*. With the ProCam-based surgical AR guidance system, direct visual guidance and augmented interaction can provide surgeons intraoperative in-situ image-guided supervision and control of robotic needle insertion.

1.4 Thesis Organization

The theme of this thesis is on investigating AR synthetic display technology and direct augmented interaction for projector-based AR system. First, ProCam system calibration and method of projection correction are proposed for construction of projector-based AR environment. Second, we propose a surface matching based registration based on ProCam system and stereovision device. Finally, we focus on the direct augmented interaction for hand gesture control and robotic needle insertion. All of them will be elucidated in the remainder of this thesis. This thesis is organized into eight chapters addressing the above questions.

Chapter 2 reviews background technologies of AR system including camera calibration, ProCam system calibration, registration methods for AR and HCI methods in surgical environment. The opportunities and challenges of dynamic ProCam calibration are also mentioned. Our contributions to ProCam calibration are further elucidated in Chapter 3. A hybrid algorithm combining extended Kalman filter (EKF) and minimal energy estimation is proposed. This is probably the first method that can achieve real-time ProCam calibration without pre-

generating structured patterns in the source image sequence. Chapter 4 reports our progress on geometric and radiometric projection correction on arbitrary surfaces. Two ProCam pixel mapping algorithms are developed to rectify projection image based on piecewise pixel matching algorithm. Chapter 5 presents our achievements on registration method for SAR. A surface matching method is proposed to register preplanned surgical models to their corresponding positions on the patient surface. Chapter 6 is devoted to augmented interaction in a surgical AR system. Methods are proposed for user-ProCam system interaction focusing on hand gesture recognition and visualization of robotic AR interaction. Our experiments and results are summarized and discussed in Chapter 7. We conclude this thesis and propose future work in Chapter 8.

Chapter 2

LITERATURE REVIEW

The theme of this thesis is on projection-based AR display and augmented interaction. We review the calibration methods for camera and ProCam system, image registration methods, and registration between the virtual scene and actual world as well as its underlying mechanisms for IGS. The opportunities and challenges of using HCI methods for augmented interaction are reviewed in the last part of this chapter. The surgical robotic system developed for supervised RFA surgery is also investigated.

2.1 ProCam System

ProCam system consists of three components: projector(s), camera(s) and a computer workstation. Most of current ProCam systems are used for spatial measurement (Figure 2-1a), projection rectification in visual-aided presentation (Figure 2-1b), and interactive display in education and entertainment.

By projecting a well-defined sequence of fringe images onto an object that is

observed by one or multiple cameras, ProCam system is used as a projection-based optical 3D scanner (Brauer-Burchardt et al., 2011)(Wladyslaw and Artur, 2009)(Fujigaki and Morimoto, 2008). Structured-intensity or -color patterns are used in the fringe images to construct the pixel correspondence between the projector and camera images. However, the structured-intensity may cause the problems of limited resolution on an irregular surface. Besides, a large number of sequential patterns are needed to be produced. For structured-color patterns, noise sensitivity might be a significant problem when they are projected onto a color-textured surface.

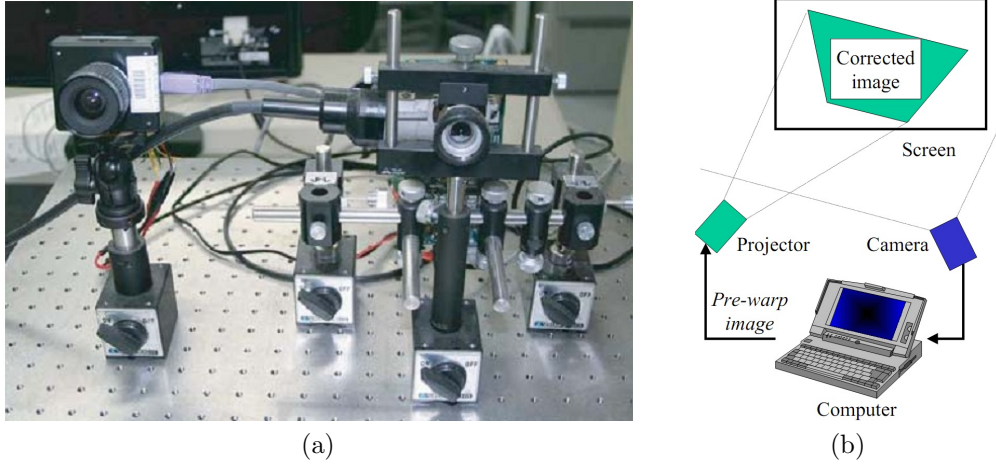


Figure 2-1. ProCam system (a) High-speed ProCam system for 3D measurement (Toma, 2010). (b) ProCam system with LCD projector and digital camera for keystone correction on the presentation screen (Sukthankar et al., 2000).

A projector with an embedded camera was developed in a ProCam system used to detect and correct the keystone distortion on the presentation screen (Li and Sezan, 2004)(Sukthankar et al., 2000). This system could be used to resolve the problems of keystone distortion on a planar surface when the projector's optical axis is not orthogonal with respect to the projection plane. However, their methods do not work for projection correction on a non-planar surface. Current work for projection correction on a dynamic non-planar surface may need an additional infrared projector to project near-infrared patterns or require a sophis-

ticated control of camera shuttering to detect the projected short-term patterns (Park et al., 2008) (Bimber et al., 2005a). With vision-based support (e.g. tracking, recognition) by camera in the ProCam system, shadows or visual markers based methods could be used for projection-based interactive display (Park and Kim, 2010). However, use of the object shadows as guidance may cause unstable tracking and non-intuitive interaction. In this study, we introduce a new integrated ProCam system to construct a surgical AR environment on an arbitrary surface (patient’s skin surface) and enable users (surgeons) to directly interact with the AR environment.

2.2 ProCam Calibration

2.2.1 Camera and Projector Calibration

The objective of ProCam system calibration is to find pixel correspondence between the projector and camera as well as their intrinsic and extrinsic parameters. In the ProCam system, cameras are responsible for acquiring the geometric information of 3D object and generating calibration parameters for the ProCam system. Camera calibration includes two phases. First, camera modelling deals with mathematical approximation of the physical and optical behavior of sensors by using a set of parameters. The second phase of camera calibration deals with estimation of the parameters, intrinsic and extrinsic parameters, with the direct or iterative methods (Hartley and Zisserman, 2003)(Turcco and Verri, 1998).

Camera modelling is mathematical projection approximation from the 3D world coordinate system to a camera coordinate system and then to a 2D camera image plane. The spatial relationship between a point in the world coordinate and its corresponding point in the image coordinate system is illustrated by the

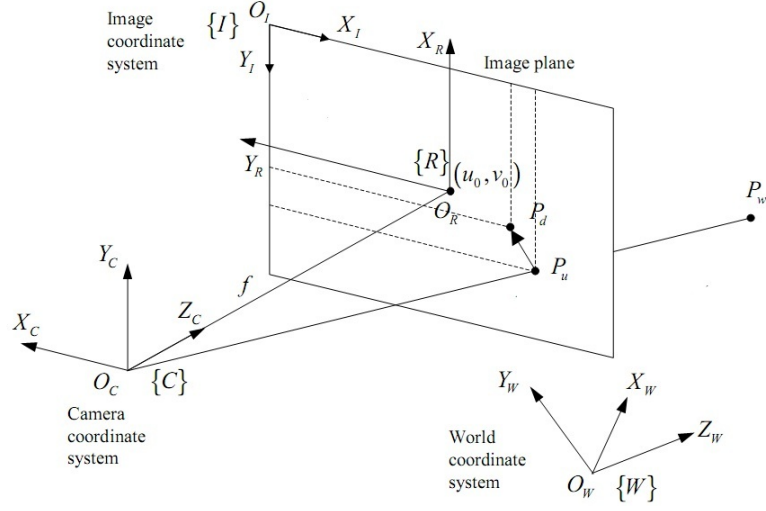


Figure 2-2. Geometric relationship between the 3D object point P_w in the world coordinate system and its corresponding point P_d in the image coordinate system (Salvi et al., 2002).

geometric model of camera imaging in Figure 2-2 (Salvi et al., 2002). In order to mitigate the lens radial and tangential distortion effects (Bradski and Kaehler, 2008), the lens distortion modelling is required to rectify the projection point P_u and then transform P_u to its corresponding undistorted point P_d (Figure 2-2). Distortion coefficients are used to construct the lens distortion models (Bradski and Kaehler, 2008):

$$\hat{x} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \quad (2.1)$$

$$\hat{y} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6). \quad (2.2)$$

$$\hat{x} = x + [2p_1 y + p_2(r^2 + 2x^2)], \quad (2.3)$$

$$\hat{y} = y + [p_1(r^2 + 2y^2) + 2p_2 x], \quad (2.4)$$

where (x, y) is the original coordinate on the image that suffers radial and tangential distortion. (\hat{x}, \hat{y}) is the corrected coordinate after correction. k_1 , k_2 and k_3 are the radial distortion coefficients, p_1 and p_2 are the tangential distortion coefficients.

The methods used for camera calibration regarding the parameters of the cam-

era models can be classified as traditional calibration methods, self-calibration, calibration based on active vision and other various calibration methods based on graphic template as well as neural networks (Wang et al., 2010b).

Traditional camera calibration method is known as using a structured calibration gadget (e.g. checkerboard) as a space reference to establish the constraints of the camera model parameters. Featured space points and image points are used as point-correspondence to construct the spatial relationship. Optimization algorithms (Unal et al., 2007) are then used to obtain the intrinsic and extrinsic parameters. With the traditional methods, most of camera models can be applied, and high precision of calibration can be achieved. The typical representatives include the direct linear transformation (DLT) methods (Li and Wang, 2007), nonlinear optimization methods (Unal et al., 2007) and two-step methods (Wu et al., 2007). The linear models, for example in the method developed by FaugerasToscani (Faugeras and Toscani, 1986), use a least-squares technique to obtain the parameters of the model. In the non-linear calibration methods, two-stage techniques are involved: firstly carrying out a linear approximation to obtain an initial guess and then a further iterative algorithm is used to optimize the parameters (Weng et al., 1992) (Wu et al., 2007). The camera calibration method using neural networks provides another solution to the camera's nonlinear models. However, this may result in large errors since the computation results may fall into a local optimal solution (Hu et al., 2007).

Besides the traditional methods, camera self-calibration method is brought into the situations where the cameras cannot be calibrated by choosing an appropriate calibration object (Wang et al., 2009). It does not depend on the calibration reference objects, the scenes and camera movements, but the self-constraints of camera intrinsic parameters and the good initial estimates.

Based on the camera self-calibration, the active calibration method is developed by controlling camera movement to overcome the cumbersome process of traditional calibration (Zhang, 2000)(Drarni et al., 2012). The active vision system enables camera installed on a controllable platform and actively controlled to obtain multiple images. The camera intrinsic and extrinsic reference parameters are then determined by using the images and the controllable camera motion parameters.

For projector calibration, the projection model can be considered as a pin-hole camera with inverse perspective projection geometry (Audet and Okutomi, 2009). In this case, the only difference between the camera and projector is the direction of the projection direction: 3D scene is projected onto the 2D image plane of the camera, or a 2D pattern of projector image is projected onto a 3D surface. The underlying mechanism of projector calibration of the model parameters is similar to camera calibration. In order to simplify the projector calibration process, a planar surface could be used as a projection surface onto which a known checkerboard or codified patterns are projected to estimate the parameters of the projector's geometric model including extrinsic and intrinsic parameters (Lanman and Taubin, 2009).

2.2.2 System Calibration

Calibration of ProCam system aims to find an accurate projector-camera pixel correspondence which could be used in projection correction including geometric and radiometric correction, and projection-object registration. Many ProCam calibration methods exist, however, most of them cannot satisfy the following two expectations: accurate pixel correspondence between the projector and camera, and real-time update of the pixel correspondence for a dynamic irregular surface.

Several methods (Falcao et al., 2009)(Sadlo et al., 2005)(Lanman and Taubin 2009) have been proposed to use a pre-calibrated projector to project feature-embedded patterns assigning pixel correspondence between the projector and camera. These methods are easy to perform, but their results may not be robust due to the projector parameters depending on the results of camera calibration. Thus, these methods are not suitable for calibration of a medical AR guidance system. A common procedure for ProCam calibration (Lanman and Taubin, 2009) is to find the homography transformation between a calibration plane and the projector image plane. However, non-linear distortion introduced by projector lenses is difficult to be modeled. Zhang and Huang (Zhang and Huang, 2006) employed checkerboard patterns instead of computing projector's corresponding points from the camera's images with structured illumination. They created new synthetic images from the projector's viewpoint and fed them to standard camera calibration tools. Resolution of the synthetic projector images that were created by an intermediate step might be low and thus lost important pixel correspondence information (Moreno and Taubin, 2012).

Currently, structured-light patterns are mostly used for calibration of the ProCam system. Sequential structured patterns are encoded and decoded by the projector and camera respectively to establish the pixel correspondence between the projector and camera. The methods of encoding the structured patterns can be classified according to different coding strategies: time-multiplexing, neighborhood codification and direct codification (Salvi et al., 2004). The advantages of time-multiplexing are easy implementation, high spatial resolution of the projection on the object surface and accurate ProCam pixel correspondence. For example, binary coded patterns are reliable and less sensitive to the surface characteristics. But it usually needs a large number of patterns projected because N patterns can only code 2^N stripes (Figure 2-3a) (Geng, 2011). The stripe patterns

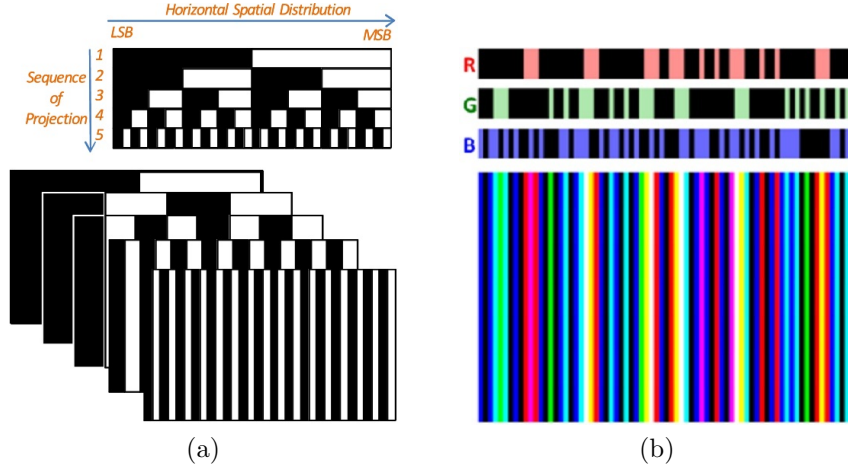


Figure 2-3. Sequential binary-coded pattern projection (a) and color stripe indexing (b) used to establish pixel correspondence in ProCam calibration (Geng, 2011).

encoded with gray code can obtain good accuracy, but the maximum resolution may not be achieved. Spatial neighborhood coding has advantages in measuring moving surfaces. However, since its codification must be condensed within a unique pattern, the spatial resolution is lower. Moreover, the measuring surface is assumed to be with locally smooth in order to correctly decode the pixel neighborhoods (Salvi et al., 2004).

The methods based on a unique pattern of a De Bruijn sequence (Figure 2-3b) have a trade off between the number of colors involved and accuracy of the acquired ProCam pixel correspondence. Most of these methods use either horizontal or vertical windows with a limited size in order to preserve the assumption of local smoothness of the measuring surface. The number of color used increases as the window size is extended to achieve a good resolution. However, more color used may increase the noise sensitivity when measuring a color textured surface (Salvi et al., 2004). Additionally, with black-white patterns replaced by a color pattern, a color camera is mandatory and camera's color calibration is also a challenge. According to Salvi et al.'s survey, direct coding method is useful to achieve large spatial resolution and few projection patterns. But this method is not suitable for

a liquid crystal display (LCD) projector based ProCam calibration and also cannot measure moving scenes since it needs additional patterns to normalize intensity or color.

Another ProCam calibration method (Chen et al., 2009) claimed to achieve an accurate result by projecting patterns on "a flat aluminum board mounted on a high precision moving mechanism". However, use of the special equipment in their system is not common and may not be applied to some surgical environments. For example, it is not allowed in a MRI-based IGS.

There are few reports on ProCam calibration for a dynamic object surface. In the study (Yasumuro et al., 2005), an additional infrared projector was employed to project near-infrared pattern projection for scanning and updating the geometry data of the object surface. Although the near-infrared patterns were imperceptible (invisible to human eyes) for the augment images, the system constitution was extended and its processing speed might be decelerated. Another hardware-controlled imperceptible pattern was generated by a digital light processing (DLP) projector with a photo transistor and a digital oscilloscope (Cotting et al., 2004). It adjusted the mirror flips to be aligned to the binary pattern for approximating RGB values. The similar application for an instantaneous adaptive display was also claimed (Cotting et al., 2005). However, this method may require sophisticated control of camera shuttering in detecting short-term patterns.

2.3 Projection Correction

In the construction of a ProCam-based AR environment, the projection usually suffers geometric distortion due to inconsistency between the projector's optical axis and user's viewing direction, or significant irregular variance of the object

surface. The radiometric distortion could be caused by the surface texture, reflectance as well as environment lighting. Since current methods for projection correction are mostly used for education, media presence and entertainment, the requirements on correction accuracy are not as strict as that of a surgical environment. The evaluation of these methods are not based on objective analysis but user-based subjective measurement (Bimber et al., 2005b)(Majumder and Brown, 2007).

Current methods for geometric correction fall into two categories regarding the priori information of the object surface: parametric and non-parametric surface (Majumder and Brown, 2007). For projection correction on the parametric surface, it is assumed that surface's geometric parameters have a direct spatial relationship with the projection image. The correspondence between the surface points and image pixels can be acquired in advance. With the mapping between the coordinates of the surface points and the camera image points, $C_{(x,y) \rightarrow (u,v)}$, and the pixel correspondence between the projector and the camera, $P_{(u,v) \rightarrow (Px,Py)}$, the final mapping between the projector image and its projection on the parametric surface could be expressed as $Dir_{(x,y) \rightarrow (Px,Py)} = C_{(x,y) \rightarrow (u,v)} \cdot P_{(u,v) \rightarrow (Px,Py)}$. The mapping $Dir_{(x,y) \rightarrow (Px,Py)}$ can then provide the mapping between the projector coordinates and the parametric surface coordinates.

While parametric methods are mostly used for planar display surfaces, they can be applied to parameterizable surfaces such as cylindrical or spherical objects. However, this method can only be used on the object surfaces with known geometric parameters. This condition is not easily satisfied in practice, because the surface parameters of the most objects are difficult to be acquired in advance.

Another method (Sukthankar et al., 2000) is widely used for rectifying the keystone distortion on the presentation screen. By identifying the quadrilateral

correspondence to the boundaries of the projection screen in the camera image, a homography transformation is used to produce a pre-warped source projector image. The keystone distortion of the projection image could be corrected with the pre-warped image projected on the presentation screen and transformed by projection transformation. This method achieves good geometric correction when the projector's optical axis is not orthogonal with respect to the presentation screen. But its use is limited to the projection correction on a planar surface

Construction of a curved projection display is significantly challenged by properly aligning the projection with surface variation. In the study (Bimber et al., 2005b), Bimber et al. developed a color structured pattern with a 2D lookup table that maps every pixel from the camera space to projector space and vice versa for view-dependent projection correction. The piecewise mapping based method achieved an obvious geometric and radiometric correction, but the color-based structured patterns may cause pixel mismatching on a color textured surface. Park et al. (Park et al., 2008) proposed a method with special embedded patterns for dynamic projection correction. In this method, each frame of the source image sequence has to be pre-processed with the special patterns to offset the alternative increase and decrease in the pixel value of the pattern images. Moreno and Taubin (Moreno and Taubin, 2012) proposed a pattern-less method of using underlying features in AR images to estimate geometric relationship between the projector and camera. It resolved the problems of intrusiveness of pattern image, however, textureless AR images made it difficult to compute the geometric relationship. Moreover, the feature extraction and matching procedure are sensitive to the noise and geometric transformation.

Radiometric distortion that is caused by surface reflectance and the environment lightings could be compensated based on the ProCam radiometric model

described in (Nayar et al., 2003) and (Majumder and Brown, 2007). Nayar, et al. (Nayar et al., 2003) proposed a radiometric model for radiometric compensation on nonspecific screens. This method adopted linear color separation and mixing for each channel to generate non-linear response. However, it is limited to a flat object surface. In another method (Majumder et al., 2003), the brightness uniformity of the projector image is achieved by matching the brightness of each pixel into a spatially-uniform dynamic range. This may cause significant loss of photometric quality of the projection image. Fujii et al. (Fujii et al., 2005) proposed a coaxial ProCam system with fixed correspondence and real-time photometric adaptation. By measuring errors between the desired and captured images, their corresponding pixel values could be updated. However, the system settings of projector and cameras may not be coaxial in a complicated surgical AR environment where a stereo camera system might be used for tracking during the guidance process. In this case, the surface reflectance has significant effects on the radiometric model.

2.4 Registration in Augmented Reality Surgery

One of the major technological issues in establishing an AR system is how to accurately register a virtual computer-generated 3D graphic model with a real scene, which means proper alignment of virtual objects with the real environment in 3D space. With this alignment, the virtual objects should behave as their real counterparts would do. Registration is a challenging task in surgical AR guidance and its requirements are difficult to satisfy because of the following two reasons.

Firstly, although there have been many effective methods and algorithms for image registration, most of them are based on image features or properties, such as image intensity (e.g. pixel-based method, cross correlation) (Antoine Maintz and Vierger, 1998)(Chmielewski and Kozinska, 2003), control points in the image

(feature-based geometric transformation) (Wen, 2008)(Wyawahare et al., 2009), frequency domain (e.g. Fourier transformation) (Manjusha and Bhosle, 2011) and entropy (multimodal registration) (Cahill et al., 2006)(Liu et al., 2002). However, registration in surgical AR guidance is not only limited to the image registration, but relies more on registration among surgical models, patient body and image slices as well as surgical tools.



Figure 2-4. Registration with clip-on marker set and reference frame for MR imager based AR surgical guidance in a biopsy experiment. (Wacker et al., 2006).

Secondly, current registration methods in IGS including VR or AR based IGS are mostly based on the manual registration which relies on surgeon's hand-eye coordination under medical images' guidance (Fichtinger et al., 2005) (Kahn and Busse, 2012). The physical sensor based registration has been used in in-situ surgical AR-guided needle biopsy (Wacker et al., 2006) (Nicolau et al., 2011). However, display area and position of surgical model is limited to the position of display device. Surgical tool's motion might be obstructed by the reference frames and limited by the indirect real-scene perception. In addition, surgeons may suffer the heavy HMD device that is attached by multiple sensors and cables (Figure 2-4).

In medical AR-guided surgery, medical images obtained from various modal-

ities usually provide valuable anatomical or functional information. The process of image-to-patient registration allows the medical images of a patient to be integrated with the patient in the physical space (the real world). Taking advantage of preoperative and intraoperative medical images, doctors could have more useful information for clinical diagnosis or further treatment.

2.4.1 AR Registration

Current registration methods for AR systems can be divided into two types: sensor based registration and computer vision based registration (Ong et al., 2004). Conventional sensor-based registration methods combine real and virtual objects using not only the cameras, but also special sensors to determine the relationships among the viewer, display device, virtual and real environments. For example, magnetic field detectors can be used to measure the positions of different objects involved in the whole augmentation process. This type of registration methods often requires bulky equipments which limit users' motion and operation (Satoh et al., 2001). The common drawback to almost all the sensors is that users must be attached to the equipments at all times.

The other type of registration, the computer vision based method, offers a potential solution for direct registration without additional sensors. The relationship among the camera, target objects and the real world is computed directly from the image sequences. The main issue in computer vision based methods is how to firstly estimate the motion and spatial structures by analyzing the image sequences and then to determine the relationships among the camera position, the real world and the virtual computer-generated objects in the same frame (Hoff et al., 1996).

In the literature, there are some reported computer vision based registration

methods (Yuan et al., 2005) (Guan and Wang, 2009). Hoff and Nguyen (Hoff et al., 1996) proposed a computer vision based registration method using a set of pre-defined passive fiducial targets on actual objects where the virtual objects will be superimposed. The main limitation of this method is it needs to keep the fiducial within the view of the head-mounted camera so that the virtual objects could be superimposed on the positions with respect to the fiducial positions. Generally, vision-based registration approaches for AR are subjected to the challenging conditions, such as wide-area work scene, changes in illumination, irregular motion and occlusion.

Simultaneous localization and mapping (SLAM), and online structure from motion (SFM) are two kinds of prevalent techniques that have been used to achieve wide-area registration for AR (Davison et al., 2007). The first real-time monocular vision-based SLAM system was implemented by Davison (Davison et al., 2003). This implementation was based on extended Kalman filter which updated each frame at cost, and it mandated the use of sparse and relatively small maps. Further, tracking was based on normalized cross-correlation (NCC) which could be disturbed easily by view-angle change and motion blur. There have been researches to improve both of the scalability and robustness of monocular SLAM. Chekhlov et al. (Chekhlov et al., 2007) improved the reliability of feature matching by employing the multi-resolution SIFT descriptor, which allowed for more agile tracking.

In most AR applications, camera intrinsic parameters are assumed to be known in advance. If the camera intrinsic parameters are changing during the augmentation process, real-time self-calibration techniques must be applied in the AR systems. However, this is not a stable and practical procedure. There are some reported calibration-free AR methods that do not require camera calibra-

tion (Hoff et al., 1996) (Kutulakos, 1998). Kutulakos (Kutulakos, 1998) proposed a calibration-free AR method based on affine object representation. However, their method could not be applied directly to the augmentation of general video sequences due to the assumption of orthographic projection.

Registration error reduction is another important issue in the development of an AR system. Registration errors result in visual conflicts between the real and virtual objects. Registration errors are difficult to control because of the high accuracy requirements and other numerous sources of errors (Harders et al., 2009). These sources of errors consist of two types, namely static errors and dynamic errors.

Static registration errors occur when relationship between the user's viewpoint and the objects in the environment remain fixed. AR applications in static conditions include fitting of assembly parts and part inspection. Possible sources of errors include optical distortion, incorrect camera focus, errors in the tracking system, mechanical misalignments and errors in computing the perspective viewing parameters (Malik et al., 2002). Dynamic registration errors occur when the user's viewpoint or the position of the objects in the environment changes. Applications in dynamic conditions include testing the functionality of a machine with moving parts to avoid collision, and changing the user's viewpoint, which is necessary in real-time repetitive registration (Bajura and Neumann, 2005). At the current stage, the registration may only work on the objects with markers. Thus, more efforts are needed to increase robustness of AR registration.

2.4.2 Registration in Image-guided Surgery

Registration of pre- and intra-interventional data is one of the key technologies in image-guided MIS, interventional radiology and endoscopy. Medical imaging

(MRI, CT) is a common preoperatively diagnostic tool, however, the information contained in these medical images cannot be directly used in surgery. For example, the medical image display may not be consistent with surgeon's viewing position during operational process. In this case, the target correspondences between the medical images and patient body have to be mentally related, which is a cumbersome process for surgeons (Chaoui et al., 2011). The approach of medical AR image registration seeks to compute these correspondences automatically and then directly augment the real surgical field with the information contained in medical images. Such an environment could provide surgeons with an enhanced visual ability to preplan, intraoperative supervise or navigate the robotic implementation throughout the whole surgical process.

To register preoperative CT images with a patient body, Fichtinger et al. (Fichtinger et al., 2005) recently developed a system for conventional CT-scanner to assist needle placement. It displays one-axial CT slice in the surgeon's field of view thereby providing a 2D internal view of the patient. Although the system is rather easy to be set up in the operating room, its application to the liver is limited because users have to reach the target with a path that belongs to the displayed slice. Thus, this can result in a suboptimal path. Moreover, some areas in the liver cannot be reached without crossing a critical structure since the needle has to remain in the displayed axial slice.

In the context of MRI-guided needle punctures, Vogt et al. (Vogt et al., 2004) displays the target defined in a preoperative MRI image and the needle position in an HMD at the same time. They reported a targeting accuracy of 1 cm on a living swine (against 3 mm on a static phantom), which was achieved by use of a rigid registration (to relate HMD and MRI frames) despite the breathing motion. Currently, very few systems designed for AR surgical guidance have been

evaluated on the real patients. This is not surprising since such an evaluation is very hard to set up. The alternative way is to elaborate a passive evaluation protocol riskless for the patient, which is hardly available in standard clinical conditions (Nicolau et al., 2009). In another study (Patriciu et al., 2005), only the robotic system has been tested on patients to evaluate its benefits in clinical conditions. Unfortunately, the resultant accuracy that can be reached in their conditions was not measured. Although a clinical benefit evaluation is mandatory, it is crucial to know the minimal system accuracy available in clinical conditions. It allows users to evaluate the risk of their movements during insertion when the needle is close to critical structures.

In another surgical robot involved experiment (Yang et al., 2010), we have developed a guidance system for thermal ablation of liver tumors with guidance of preoperative CT images. The system was designed in order to assist radiologists when they choose the needle path toward the target. A static abdominal phantom was showed that the targeting accuracy was about 2.5 mm. Since the patient is not static in practice, simulation of organ motion with preoperative 3D CT or MRI images (Nicolau et al., 2009) could be used to predict the tumor position during intraoperative intervention. This reasonable assumption is based on the clinical studies of organ repositioning error of a patient, under the breathing monitoring, which is evaluated below 2 mm.

Currently, few reports have been published on surgical guidance with projector-based AR. Gavaghan et al. (Gavaghan et al., 2011) described a miniature laser projection-based image overlay system that was established on a hand-held laser projector with passive markers and an optical tracking system. Unlike conventional projectors, their system incorporated a Microvision Development Kit (PicoP, Microvision Inc., WA) containing a portable RGB laser projector, a video

processor and a microelectromechanical system (MEMS) controller. An additional navigated calibration plate has to be used in their calibration procedure. Thus, with this projector system, surgeons may manipulate less surgical tools in hands or have collaborative problems with their assistants during the surgery. In the study conducted by Krempien (Krempien et al., 2008), projection-based visual guidance was experimented for interstitial brachytherapy. The expected projection position of tumor model was determined by real-time CT scanning of patient body and model-based surface matching. However, high cost, long scan time and hazardous radiation from X-rays may prohibit the surgeon from continuously performing such operation.

2.5 Human-computer Interaction in VR and AR Environment

HCI design aims to enable the interaction between users and computer-based environment more controllable, friendly and robust. Two main aspects should be considered when a HCI system is designed: functionality and usability. Interaction between AR and users is a complex subject which is linked to a broader but strongly related subject, called Natural User Interfaces (NUI) (de Landgraaf, 2009). This paradigm is based on the assumption that users should interact with the virtual world just as they would with the real world.

HCI design may play an important role in the minimally invasive surgery with image guidance. In the intraoperative procedure, some characteristics must be considered when building a new human-computer interface in order to obtain medical approval (Sielhorst et al., 2008): (1) sterilization of the complete operating room (the entire environment and all the equipments); (2) no discharged air of

computers or projectors; (3) Handy devices with as few cables as possible; and (4) system interaction ergonomics. Patient safety and operation accuracy can benefit from a user interface with intraoperative visual-guidance information and non-contact interaction.

2.5.1 HCI Design and Methods

In HCI design, degree of the activity that involves a user with a machine should be considered. The user activity has three different levels (Karray et al., 2008): physical, cognitive and affective. The physical aspect determines the mechanics of interaction between human and computer, while the cognitive aspect deals with ways that users understand and interact with the system. The existing physical technologies for HCI basically can be categorized by the relative human sense that the system (machine) is designed for. HCI design of the system relies on three human senses: vision, audition and touch. Currently, many experimental interfaces have also been proposed based on body gestures, speech and brain computing (Teeni, 2006).

In VR system (Iwata, 2003), users could use the specialized equipment and devices, for example, data glove is used to track hand motion, interpret hand gestures and provide system input. Interactive 3D computer graphics and audio software generate the appropriate scenes and auditory cues. However, these equipments may have disadvantages such as high cost to use and unnatural interaction.

Besides the specialized devices or equipments developed for human gesture recognition, vision-based hand gesture interaction is widely used which has better ergonomics for direct HCI. Image processing and segmentation algorithms are performed to capture shapes of hand and extract gesture features. A modified direction analysis algorithm (Khan and Ibraheem, 2012) was developed to find

the statistical parameters (variance and covariance) to compute slope and trend of the hand plane. Another method (Hasan and Mishra, 2010) applied multivariate Gaussian distribution to recognize hand gestures using non-geometric features. The input hand images are segmented using two different methods: skin color based segmentation with HSV color model and clustering based thresholding techniques.

Neural network is another important method used to recognize human gestures (Khan and Ibraheem, 2012). In the recognition method developed by Kulkarni and Lokhande (Kulkarni and Lokhande, 2010), the input images are firstly converted into HSV (Hue, Saturation, Brightness) color model for hand (region) segmentation. The gesture features, extracted using histogram technique and Hough algorithm, are then classified by feed forward Neural Networks (Figure 2-5). The classification phase includes network architecture design, network creation and network training. The network is trained on 8 samples of each sign. In their experiments, samples of same size and other features like distance, rotation and lighting effect with uniform background are taken into consideration. However, this method does not work well for all the 26 signs because the recognition rate highly depends upon the extracted gestures' geometric features.

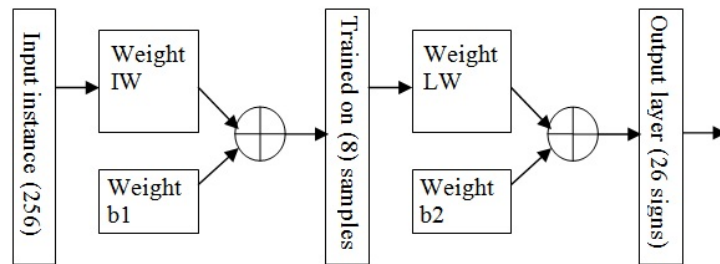


Figure 2-5. Neural network used for classification of gestures features (Kulkarni and Lokhande, 2010).

A common approach to describe the human motion is using a state-based model, such as Hidden Markov Model (HMM), to convert a series of motions into

description of activity. Due to having advantages in clustering spatio-temporal patterns, HMM has been widely used in speech and handwriting recognition. An early effort by Yamato et al. (Yamato et al., 1992) used discrete HMM to recognize image sequences of six different tennis strokes among three subjects. Liu et al. (Liu et al., 2004) designed a letter gesture input system to classify HMM model training methods. The system acquired 25 frames of video data and used skin color segmentation to locate the hand. Along each trajectory, orientation of each of the 25 hand movements was computed and quantized to one of 18 discrete symbols. This discrete observation sequence was fed into an HMM classification module for training and testing. Popularity of HMM is largely due to their ability to learn model parameters from observation sequence through Baum Welch (Yamato et al., 1992) and other re-estimation procedures.

2.5.2 Augmented Interaction

As introduced in Chapter 1, augmented interaction is a new style of human-computer interface that tries to make users interact with the real and virtual environment simultaneously. The computer's role is not an outstanding "control panel" but an intelligent controller running behind user interface to assist the augmented interaction. The user's situation can be automatically recognized by applying a range of recognition methods allowing the computer to understand users' inputs and responses with AR display. With this style, a user can focus on AR objects as well as the surrounding real environment which is augmented by computer-generated synthetic information.

Currently, some trials on augmented interaction have been conducted for VR-based design. Stark et al. (Stark et al., 2010) developed a hybrid immersive modelling environment by merging desktop computer-aided-design (CAD) with

VR technologies for art design. Digital media has advantages of allowing integration of a developed product model with concept design and other associated properties which are easily revised. In addition, some downstream processes, such as planning and fabrication, can be integrated (Nee et al., 2012). Although they can enhance efficiency product design, there is a lack of direct and intuitive interaction.

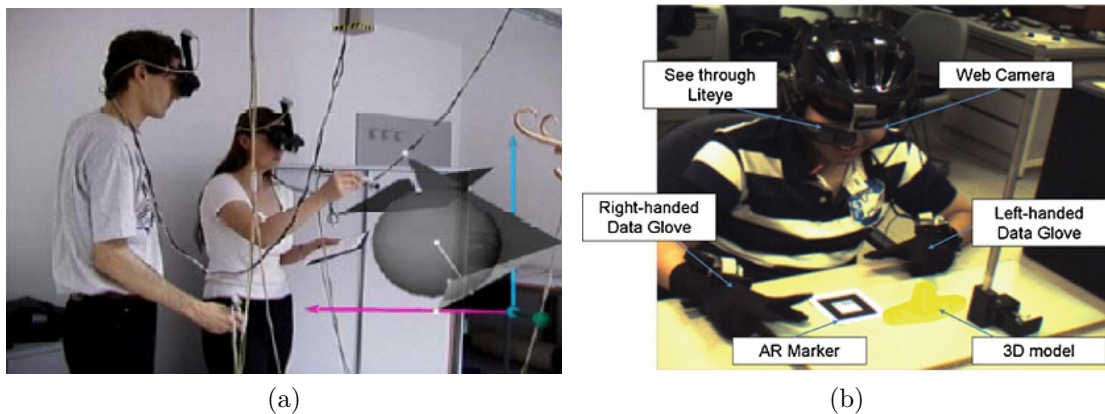


Figure 2-6. Augmented interaction with attached devices and equipments (Nee et al., 2012).

AR-based systems support information visualization through augmenting the real world with computer-generated models. However, only a few systems support product creation and modification in an AR-based environment using 2D or 3D interaction tools. Applying these interaction tools usually requires users to wear or to attach additional devices, which could cause unsatisfying ergonomics and reduce the AR immersion effects (Figure 2-6).

Augmented interaction has been introduced in the prototyping process. In the automotive industry (Frund et al., 2005), AR has been used for assessing interior design by overlaying different car interiors, which is usually available as 3D models are developed to car body mock-ups in the initial phase. However, complex parts combining the primitives cannot be directly examined if the system does not support direct interaction. Park (Park, 2004) presented an AR-based reformable

mock-up system for design evaluation, in which an interactive modification of shapes, colors, textures and user interfaces can be carried out.

In order to enhance manual dexterity and accuracy of instrument manipulation in surgery, robotic assistance has been introduced in the augmented MIS to solve operation constraint problems. The da Vinci Surgical System, a master-slave integrated robot, has been applied to treat a range of surgical conditions improving operational flexibility and accuracy (Akinbiyi et al., 2006). In the sophisticated minimally invasive operations like coronary artery bypass and mitral valve repair, robot assisted implementation could achieve satisfying clinical outcomes (Yamamoto et al., 2005). Surgical robot also demonstrated its advantages in speed, accuracy and consistency in larger liver tumor treatment with RFA. Our group has investigated design and implementation of a novel robotic system for image-guided RFA in which both robotic kinematics and clinical requirements could be satisfied (Yang et al., 2010). Other advantages offered by the robotic RFA include minimizing radioactive exposure during CT scan and overcoming visual restriction and spatial constraint in minimally invasive procedures (e.g. laparoscopic or percutaneous surgical operation). Integration of imaging guidance with surgical robot assistance will bring the preoperative planning and intraoperative manipulation into perspective for the minimally invasive treatment.

2.6 Summary

In Chapter 2, we review all background technologies in constructing a ProCam-based surgical AR guidance system. It includes camera calibration and ProCam system calibration, registration among projection image, surgical model and the object surface (patient's skin surface), and HCI methods used for current VR and AR interaction. As to camera and projector calibration, their underlying mech-

anisms are examined. Different structured-light methods for ProCam calibration are investigated, and their advantages and disadvantages are compared. The opportunities and challenges in projective geometric and radiometric correction are reviewed. In the second part of Chapter 2, we point out weakness of the current registration methods including image registration methods used for medical AR guidance. AR registration methods are investigated in accordance with sensor based and computer vision based registration. Human-computer interaction plays an important role in an interactive surgical AR guidance system. We investigated the opportunities of developing a hand gesture based HCI for direct augmented interaction. Integration of augmented interaction with surgical robot implementation is discussed in Section 2.5.

Chapter 3

SYSTEM CALIBRATION

System calibration encompasses monocular and binocular camera calibration, and calibration of the ProCam subsystem which establishes optimal projector-camera pixel correspondence with respect to the current patient (skin) surface. In Section 3.1, we describe the geometric models of camera and projector calibration. The general pinhole camera model presented in Section 2.1 is extended with lens distortion. The calibration results, which are obtained from camera calibration in Section 3.2, are used for Section 3.3—ProCam system calibration. In order to mitigate projection distortion on the patient surface, a novel hybrid algorithm is presented to achieve real-time prediction of projection field to dynamically calibrate the ProCam system.

3.1 Camera and Projector Calibration

In the camera calibration, we aim to find parameters of the pin-hole model based camera imaging model (Figure 2-2). It consists of the intrinsic and extrinsic parameters including focal length, principal points, rotation matrix and translation

vector, which establish point-pixel mapping between the world and camera coordinate systems. In total, 11 parameters (5 intrinsic and 6 extrinsic parameters) are required to be estimated in the calibration procedure.

Suppose $P(X_w, Y_w, Z_w)$ is a spatial point in the world coordinate system, while $p(u, v)$ is the corresponding point in the camera coordinate system. Relationship between the point $P(X_w, Y_w, Z_w)$ and $p(u, v)$ can be modeled as

$$p = s \begin{bmatrix} f_x & a & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [\mathbf{R} \ \mathbf{T}] P, \quad (3.1)$$

where $\mathbf{R}_{3 \times 3}$ and $\mathbf{T}_{3 \times 1}$ are the extrinsic parameters denoting spatial rotation and translation matrix respectively. The composite matrix \mathbf{RT} represents the transformation from the world space to the camera coordinate system. f_x, f_y, c_x, c_y and α are the intrinsic matrix of the camera (projector) which are the focal length, principal points and skew of the image axes respectively. s is a scale factor used to adapt to a variety of uncertainties brought by sampling on the horizontal direction in the computer images (Hartley and Zisserman, 2003).

In practice, the lens distortion model mentioned in Chapter 2 must be estimated as well. In this study, calibration method introduced by Bouguet (Bouguet, 2010) is adopted to estimate radial and tangential distortion coefficients of the lens distortion. A planar checkerboard pattern is observed in multiple orientations, and corresponding feature points that locate on the checkerboard and captured images are recorded to construct the spatial relationship in accordance with the camera imaging model. In order to solve the intrinsic parameters of the model, a two-stage non-linear method is used to firstly carrying out linear approximation to obtain an initial guess, and then iterative regressions to compute the optimal parameters. A

single view of the checkerboard can then be used to estimate the extrinsic parameters of the camera imaging model. These parameters are used for the mono- and binocular camera calibration. As to calibration of the projector, mapping from the 2D images to 3D points is referred as perspective projection and the projector model is assumed to be an inverse camera pin-hole model.

3.2 Calibration for Static Surface

In order to establish the optimal ProCam pixel correspondence, structured-light illumination is used to encode and decode the projector and camera image pixels respectively. Comparing to the binary code projection, successive numbers of the gray code introduce less pixel misplacement because they vary exactly in each bit and have double width of bright and dark lines in the pattern with high resolution. This could facilitate decoding of camera images, especially at steep object surfaces where the code appears to be compressed. Figure 3-1 illustrates construction of the pixel correspondence between the projector and camera image sequences by structured-light illumination.

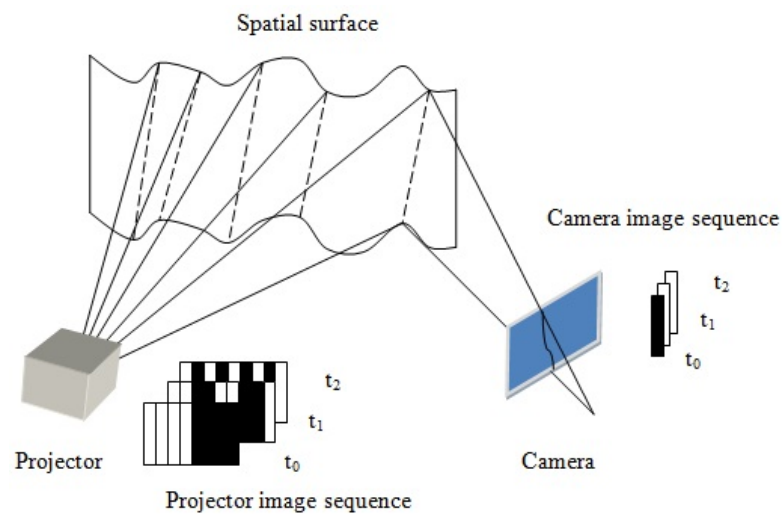


Figure 3-1. Encoding of projector images and decoding of camera images.

Since the gray code determines absolute range of phase without any ambiguity while the phase shift offers subpixel resolution, a hybrid method combining gray-coded pattern and phase shift is developed in this study to search the pixel correspondence with high accuracy. The phase is generated by the following equations that describe the intensity values of the three measured fringe patterns:

$$I_1(x, y) = I_0(x, y) + I_{mod}(x, y) \cos(\varphi(x, y) - \alpha), \quad (3.2)$$

$$I_2(x, y) = I_0(x, y) + I_{mod}(x, y) \cos(\varphi(x, y)), \quad (3.3)$$

$$I_3(x, y) = I_0(x, y) + I_{mod}(x, y) \cos(\varphi(x, y) + \alpha), \quad (3.4)$$

where I_1, I_2, I_3 are the intensity of the fringe patterns produced by the projector images varying with the phase,

$$\varphi(x, y) = \arctan\left[\sqrt{3} \frac{I_1(x, y) - I_3(x, y)}{2I_2(x, y) - I_1(x, y) - I_3(x, y)}\right], \quad (3.5)$$

and the phase shift α . I_0 is the intensity obtained by averaging the white and black illumination, and I_{mod} is the modulation amplitude.

In Equation 3.5, the phase calculation might be slow due to processing of arc-tangent function. In order to accelerate this process, a trapezoidal phase-shifting method is developed to simplify the phase unwrapping. The calculation of Equation 3.5 can be sped up by calculating an intensity ratio using a simple function instead of the arctangent function. By calculating defocused trapezoidal patterns instead of the sinusoidal patterns, calculation speed of the trapezoidal method can be preserved while achieving the same accuracy of the traditional sinusoidal phase-shifting algorithm (Huang et al., 2005). From these three intensity values, the intensity ratio $r(x, y)$ can be acquired by equation

$$r(x, y) = \frac{I_{med}(x, y) - I_{min}(x, y)}{I_{max}(x, y) - I_{min}(x, y)}, \quad (3.6)$$

where $I_{min}(x, y)$, $I_{med}(x, y)$ and $I_{max}(x, y)$ are the minimum, median and maximum intensity values in the phase-shifted sinusoidal patterns respectively. The phase can then be calculated by the following equation:

$$\varphi(x, y)' = \frac{\pi}{3} [2 \times \text{round}(\frac{N-1}{2}) + (-1)^{N-1} r(x, y)]. \quad (3.7)$$

Corresponding pixels between the projector image plane and the camera image planes are matched by decoding the captured image sequence which consists of encoded columns and rows. The decimal integer index of the corresponding projector column or row is recovered by decoding the received bit sequence on each camera pixel (Figure 3-2). In order to define the illumination areas, a user-selected intensity threshold can be used to determine whether a given pixel is illuminated.

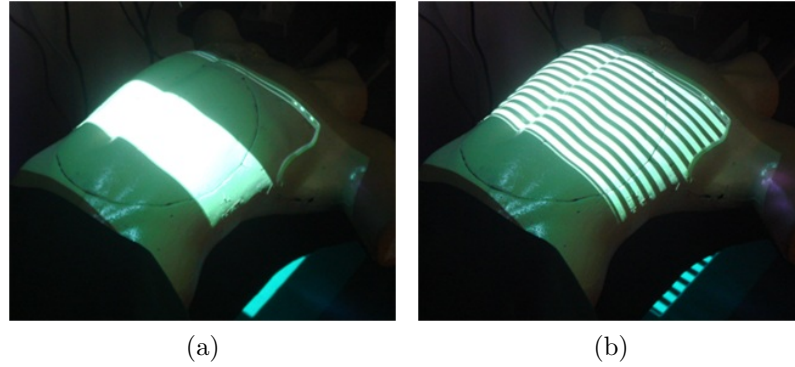


Figure 3-2. Projection of binary coded patterns with phase shift variation. (a) A binary-coded pattern of two strips. (b) A binary-coded pattern of twenty-four strips with phase shift.

3.3 Calibration for Dynamic Surface

In this section, we aim at developing an effective method to acquire the real-time pixel correspondence between the projector and camera for dynamical calibration

of the ProCam system. This dynamic ProCam pixel correspondence can be used to produce the real-time rectified overlaid images on the patient body to counteract projection distortion caused by patient free breathing or slight motion.

A hybrid non-pattern-embedded calibration algorithm is proposed which combines Kalman filter based recursive tracking and minimal energy based curvature comparison. A dynamic projection field is established in real time by tracking the initial feature points and updating their motion parameters. With the initial matching candidates, their local curvature parameters are computed and compared with the ones in the previous frame within searching windows which are defined by the motion centers of feature groups. Minimal bending energy is then adopted as an optimization criterion to find the corresponding points from the correspondence candidates. In this way, the pixel mapping between the projector and camera (ProCam mapping) is dynamically established. This hybrid algorithm combines Kalman filter based feature tracking and optimization of a bending energy model. Without using any embedded structured patterns in the projector image sequence, this algorithm achieves direct and instantaneous acquisition of dynamic ProCam pixel correspondence.

The overview of this hybrid algorithm for ProCam calibration is shown in Figure 3-3. Without using embedded structured-patterns, this algorithm uses nature surface points P_s as feature points. Initialization of the feature correspondence in the camera image I is a mapping process from the surface points P_s to the image points P_i . In order to achieve this initialization, an initial ProCam mapping \mathbf{M}_{p-co} is established in advance with the method of phase-shifting projection introduced in Section 3.2. For those P_s without corresponding points in I , nearest neighbour interpolation is used to find the corresponding transformation vectors.

With the initialized feature correspondence in the camera image, a recursive

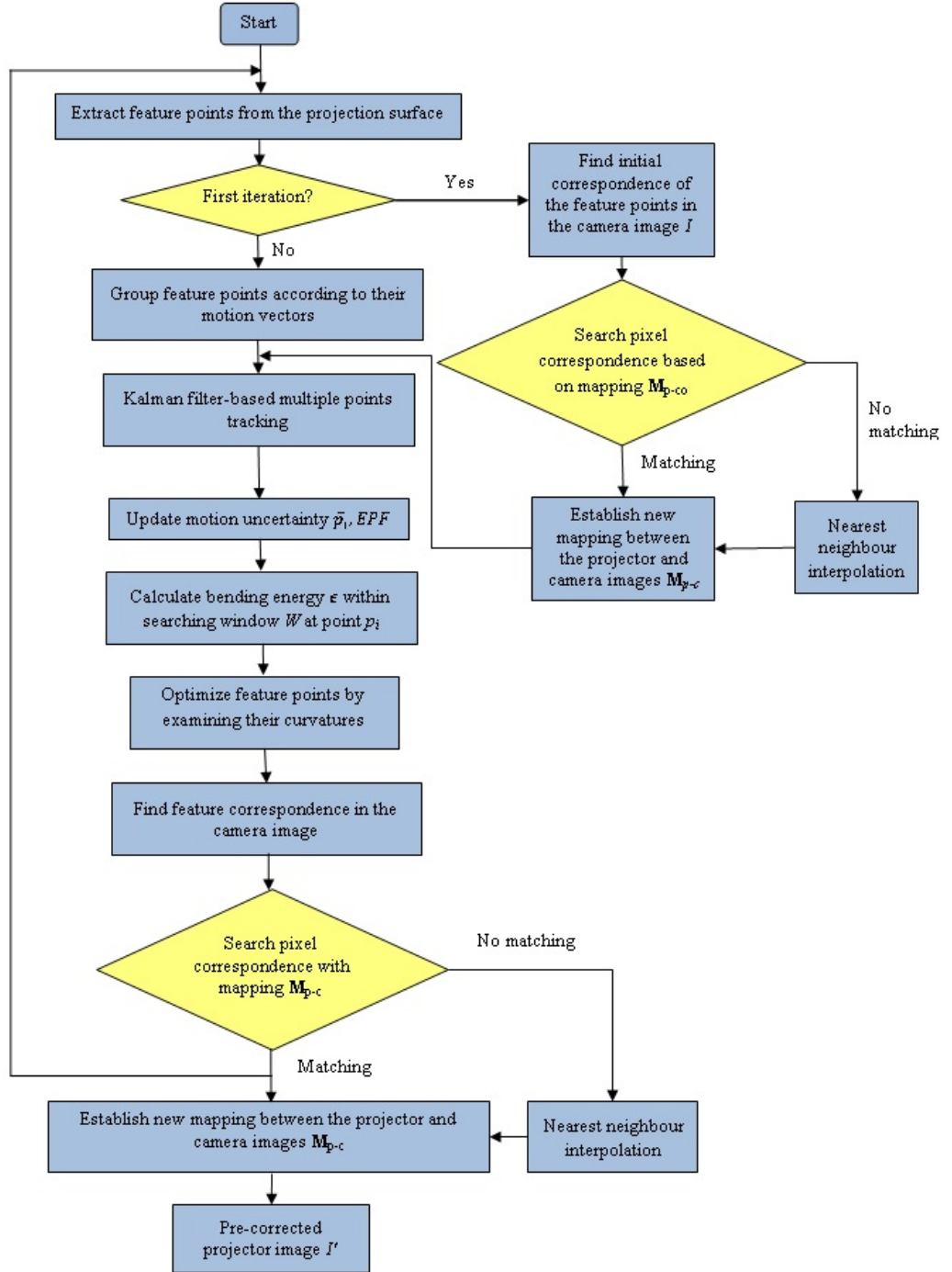


Figure 3-3. Workflow of the hybrid algorithm.

tracking process based on multi-Kalman filter approach (Gao et al., 2006) is used to update the candidates of the feature correspondence (CFC) for every captured frames. In this study, we assume the skin of human chest and belly is elastic

curved surface. Affected by different forces distributed inside the patient body, the CFC are grouped according to their average motion vectors and centers (Figure 3-7). In order to refine the candidates to find the most optimally matched feature correspondence, bending energy at the CFC in different groups is estimated and compared with the one in the previous frame. We consider the candidate features with the optimized curvature parameters (calculated by principal curvatures) as the points corresponding to the surface feature points in the current frame. The pseudo code of the hybrid algorithm including EFK tracking and bending energy minimization is presented in Algorithm 1.

The searching window W , which includes the feature candidates for current frame, is established based on the motion centers of the feature points. For those CFC without their corresponding points in the current frame, the nearest neighbour interpolation is applied. The feature correspondences in the camera image of the initial surface feature points are then updated frame by frame to generate a dynamic ProCam mapping \mathbf{M}_{p-c} based on the initial ProCam mapping \mathbf{M}_{p-co} (Figure 3-5). The details of this hybrid algorithm are presented in the following sections.

3.3.1 Feature Initialization in Camera Image

The nature surface points P_s are used in this study instead of the structured projection patterns (Park et al., 2008) to establish a dynamic ProCam pixel correspondence. This may enable the ProCam system to realize an instantaneous patternless display without additional infrared projector detecting or pattern-sequence encoding. Thus it enhances flexibility and adaptability of the ProCam-based AR display. P_s are derived from edge maps of each frame. Canny edge detector can be used to generate the edge maps from the gray or color input images. For color

Algorithm 1 Hybrid Algorithm (tracking-matching)

```
1. /*Key: Key input by users.*/
2. while (Key  $\neq$  'Esc') do
3.   if (iteration == 1)
4.     /*Compare curvature parameters between points in
5.       the two consecutive frames*/
6.     /*  $\epsilon(x)$ : local bending energy at point  $x$  */
7.     for  $i = 1$  to  $n$ , do
8.       for  $j = 1$  to  $n$ , do
9.          $T_i = \epsilon(P'_i) - \epsilon(P_j)$ 
10.         $P'_i = \text{argmin}(T_i)$ 
11.      endfor
12.    endfor
13.  else
14.    /*Group the motion vectors  $\mathbf{v}_i$ */
15.    for  $i = 1$  to  $n$ , do
16.       $\mathbf{v}_i = \mathbf{v}(P_{i(t+\Delta t)}) - \mathbf{v}(P_{it})$ 
17.      /* $\mathbf{q}_i$ : grouping criterion*/
18.      if  $\mathbf{v}_i < \mathbf{q}_i$ 
19.         $\mathbf{v}_i \in \mathbf{G}_i$  /* $\mathbf{G}_i$ : group  $i$ */
20.      endfor
21.    /* $\bar{\mathbf{v}}_i$ : average motion vectors of the group  $i$ 
22.    Er: error for prediction of  $\bar{\mathbf{v}}_i$ */
23.    for  $i = 1$  to  $n$  do
24.       $\bar{\mathbf{v}}_i = \text{average}(\mathbf{v}_s + \dots \mathbf{v}_t) | s, t \in \mathbf{G}_i$ 
25.      Er =  $E(\mathbf{v}_i, \bar{\mathbf{v}}_i)$ 
26.    endfor
27.    /*Kalman filter-based tracking */
28.    state:  $\mathbf{x}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{u}_k + \mathbf{w}_k$ 
29.    observation:  $\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k$ 
30.    /*Prediction phase*/
31.     $\bar{\mathbf{x}}_k = \mathbf{F}_k \mathbf{x}_{k-1} + \mathbf{B}_k \mathbf{u}_k$ 
32.     $\bar{\mathbf{P}}_k = \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T + \mathbf{Q}_k$ 
33.    /*Update phase*/
34.    Kalman gain:  $\mathbf{G}(\text{Er}) = \bar{\mathbf{P}}_k \mathbf{H}_k^T [\mathbf{H}_k \mathbf{P}_k \mathbf{H}_k^T + \mathbf{E}_k]^{-1}$ 
35.     $\mathbf{x}_k = \bar{\mathbf{x}}_k + \mathbf{K}[\mathbf{z}_k - \mathbf{H}_k \bar{\mathbf{x}}_k]$ 
36.     $\mathbf{P}_k = [1 - \mathbf{K} \mathbf{H}_k] \bar{\mathbf{P}}_k$ 
37.    /*Compare curvature difference*/
38.    for  $i = 1$  to  $n$ , do
39.      for  $j = 1$  to  $n$ , do
40.         $T_i = \epsilon(P'_i) - \epsilon(P_j)$ 
41.         $P'_i = \text{argmin}(T_i)$ 
42.         $\Rightarrow$  new mapping  $\mathbf{M}_{p-c} : P'_i \rightarrow P_p$ 
43.      endfor
44.    endfor
45.  endwhile
46. Return the new mapping  $\mathbf{M}_{p-c}$ 
```

images processing, the edge maps of the R, G, B channel of the region of interest (ROI) are produced respectively. An overall edge map can then be obtained by taking logic OR of the three edge maps with their respective channels R, G, B. The feature points are selected by sub-sampling the edge points from the overall edge maps. Figure 3-4 shows a projected medical image of vessel and organ as well as their overall edge maps with feature points.

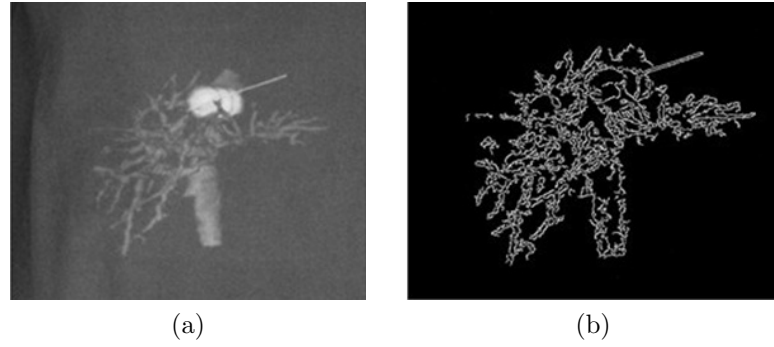


Figure 3-4. Projection image on the patient surface (a) and its corresponding edge map with surface feature points (b).

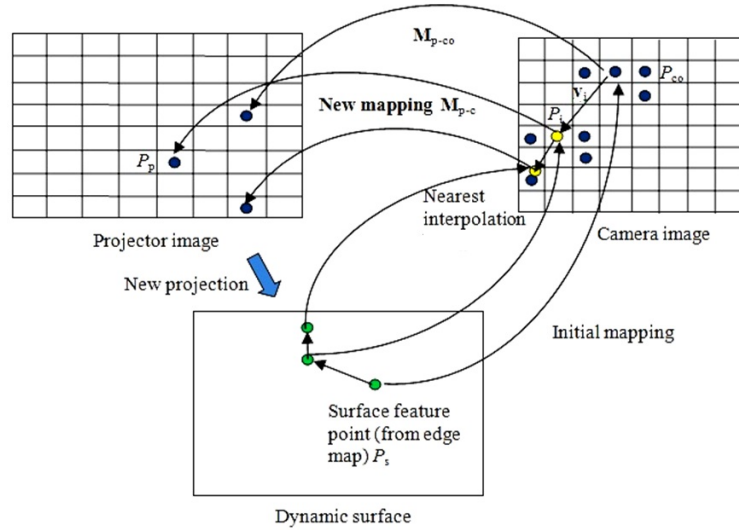


Figure 3-5. New mapping establishment for the ProCam system. M_{p-co} is the initial pixel mapping between the projector and camera image. P_{co} are the initial corresponding points of the projector image points P_p on the camera image. The surface feature points P_s (green points) have their corresponding points P_i (yellow points) on the camera image. The new corresponding points of P_i can be found by 2D lookup table from the mapping M_{p-co} . For the P_i without corresponding points on the camera image, nearest neighbour interpolation is used to find their correspondences P_{co} .

Dynamic pixel mapping between the projector and camera can be established as shown in Figure 3-5. In the initialization procedure, a group of binary encoded feature points are used to establish the initial ProCam mapping \mathbf{M}_{p-co} . A 2D lookup table is generated from \mathbf{M}_{p-co} to establish pixel correspondence between the projector and camera coordinate system. When the surface feature points (green points) are captured, their corresponding points in the camera image (yellow points) have a new mapping \mathbf{M}_{p-c} from camera to the projector image (the blue points P_p on the projector image) based on the mapping \mathbf{M}_{p-co} , $P_p \rightarrow P_{co}$. With the pixel mapping $P_s \rightarrow P_i$, $P_i \rightarrow P_{co}$ and $P'_s \rightarrow P'_i$, update of the feature points on the camera image, $P_i \rightarrow P'_i$ can then be derived. In this way, the correspondence mapping between the projector and camera image, $P'_i \rightarrow P'_p$ can be acquired. P'_s , P'_i and P'_p are the updates of the corresponding feature points on the object surface, camera and projector image respectively. Nearest neighbor interpolation is implemented for P_i if there is no mapping from P_i to P_{co} . In this way, the ProCam mapping \mathbf{M}_{p-co} can be updated to establish the real-time pixel mapping between the projector and camera images.

3.3.2 Tracking of Multiple Feature Points with Extended Kalman Filter

In order to dynamically generate ProCam mapping \mathbf{M}_{p-c} to adapt to a varying (skin) surface, shift of the corresponding feature points in the camera image need to be predicted frame by frame. In this hybrid algorithm, extended Kalman filter (EKF) is used as a vision based tracking algorithm to locate the searching window regions that include the candidate corresponding feature points. Since the skin surface motion during free breathing is continuous and smooth, the EKF, which is used to build a model that maximizes the posteriori probability of the previous

measurements to predict the parameters of nonlinear motion, is an ideal tracker for the projected feature points which varies with the patient surface motion. The following two assumptions are made for this tracking. Since shift of the projection field on the patient surface due to breathing motion is within a small area, motion of the projection point P_i on the patient body between the two consecutive frames k and $(k + 1)$ is assumed to be linear. Besides, we assume the noise measurement associated with random image noise has Gaussian distribution.

Assume spatial points $P_{k-1} = (X_{k-1}, Y_{k-1}, Z_{k-1})$ and $P_k = (X_k, Y_k, Z_k)$ are two corresponding feature points which are captured in the frame at k and $(k + \Delta k)$. Their corresponding points on the camera image plane are $p_1 = (u_{k-1}, v_{k-1})$ and $p_2 = (u_k, v_k)$. Then we have Equation 3.8 and 3.9 by homographic transformation matrix,

$$u_i = \frac{r_{11}x_i + r_{12}y_i + r_{13} + \frac{t_x}{Z_i}}{r_{31}x_i + r_{32}y_i + r_{33} + \frac{t_z}{Z_i}}, \quad (3.8)$$

$$v_i = \frac{r_{21}x_i + r_{22}y_i + r_{23} + \frac{t_y}{Z_i}}{r_{31}x_i + r_{32}y_i + r_{33} + \frac{t_z}{Z_i}}. \quad (3.9)$$

Assume the homogeneous transformation between P_{k-1} and P_k is \mathbf{H} .

$$P_{k-1} = \mathbf{H}P_k, \quad (3.10)$$

$$\begin{bmatrix} X_{k-1} \\ Y_{k-1} \\ Z_{k-1} \end{bmatrix} = \mathbf{R} \begin{bmatrix} X_k \\ Y_k \\ Z_k \end{bmatrix} + \mathbf{T}, \quad (3.11)$$

$$\mathbf{H}' \begin{bmatrix} u_{k-1} \\ v_{k-1} \\ 1 \end{bmatrix} = \mathbf{R} \left[\mathbf{H}' \begin{bmatrix} u_k \\ v_k \\ 1 \end{bmatrix} \right] + \mathbf{T}, \quad (3.12)$$

$$\mathbf{H}' \begin{bmatrix} u_{k-1} \\ v_{k-1} \\ 1 \end{bmatrix} = \mathbf{R}'' \begin{bmatrix} u_k \\ v_k \\ 1 \end{bmatrix} + \mathbf{T}'', \quad (3.13)$$

and the point (u_k, v_k) at frame k can be derived by

$$u_k = \frac{r_{11}u_{k-1} + r_{12}v_{k-1} + r_{13} + \frac{t_x}{Z_{k-1}}}{r_{31}u_{k-1} + r_{32}v_{k-1} + r_{33} + \frac{t_z}{Z_{k-1}}}, \quad (3.14)$$

$$v_k = \frac{r_{21}u_{k-1} + r_{22}v_{k-1} + r_{23} + \frac{t_y}{Z_{k-1}}}{r_{31}u_{k-1} + r_{32}v_{k-1} + r_{33} + \frac{t_z}{Z_{k-1}}}. \quad (3.15)$$

From Equation 3.12 and 3.13, we can establish mapping \mathbf{H}' and \mathbf{H}'' , $\mathbf{H}'' = \mathbf{R}''\mathbf{T}''$, between the two corresponding points in the camera images. In this case, Equation 3.11 can be written as

$$f(P_k, P_{k-1}, Z_{k-1}, \mathbf{R}_{k-1}, \mathbf{T}_{k-1}) = 0. \quad (3.16)$$

Before the feature points in each captured frame are determined, the motion vectors are grouped according to skin surface' motion patterns caused by patient free breathing (Figure 3-6). The feature points p_i with motion vectors $\mathbf{v}_i = (v_i^\alpha, v_i^\beta, v_i^\gamma)$,

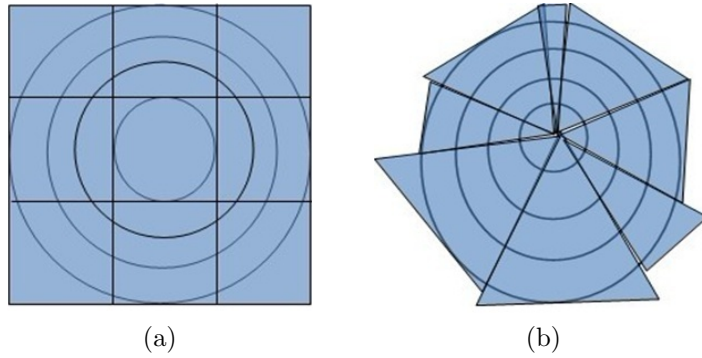


Figure 3-6. Patterns for motion vectors grouping. Motion vectors grouping (b) depends on motion motivated by the internal force under the surface (a).

$$v_i^\alpha - v_{(i\pm 1)}^\alpha < q,$$

$$v_i^\beta - v_{(i\pm 1)}^\beta < q,$$

$$v_i^\gamma - v_{(i\pm 1)}^\gamma < q,$$

are grouped together, where i is the feature number, q is the threshold for vector grouping. Groups are numbered to implement prediction of their mean motion vector, \bar{p}_{ni} , and the error of predicted projection field (EPF), \mathbf{w}_i , respectively. \bar{p}_{ni} is the motion center of feature group n at frame k . In this study, the motion vector of point p_i is defined as $\mathbf{v}_i = \mathbf{X}_k$, $\mathbf{X}_k = (\theta_x, \theta_y, \theta_z, t_x, t_y, t_z)$. The EPF is a covariance matrix,

$$\mathbf{w}_i = \begin{bmatrix} \sigma_{rx}^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma_{ry}^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{rz}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{tx}^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sigma_{ty}^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_{tz}^2 \end{bmatrix}. \quad (3.17)$$

With EPF, error prediction of the motion center $\bar{p}_{ik} = (u_i, v_i)$, \mathbf{U}_{pik} , at frame k can be obtained from,

$$\begin{aligned} \mathbf{U}_{pik} &= E(\mathbf{p}_{ik} - \bar{\mathbf{p}}_{ik})(\mathbf{p}_{ik} - \bar{\mathbf{p}}_{ik})^T \\ &= \frac{\partial \mathbf{p}_{ik}}{\partial \mathbf{X}_{ik}} \big|_{\mathbf{x}_k = \bar{\mathbf{x}}_{nk}} * \mathbf{w}_{ik} * \left[\frac{\partial \mathbf{p}_{ik}}{\partial \mathbf{X}_{ik}} \big|_{\mathbf{x}_k = \bar{\mathbf{x}}_{nk}} \right] \\ &= \frac{\partial f_i}{\partial \mathbf{X}_{ik}} \big|_{\mathbf{x}_k = \bar{\mathbf{x}}_{nk}} * \mathbf{w}_{ik} * \left[\frac{\partial f_i}{\partial \mathbf{X}_{ik}} \big|_{\mathbf{x}_k = \bar{\mathbf{x}}_{nk}} \right], \end{aligned} \quad (3.18)$$

where \bar{p}_{ik} is the average of feature point p_i belonging to the group n ,

$$\bar{p}_{ik} = \begin{bmatrix} \frac{r_{11}u_i + r_{12}v_i + r_{13} + t_x/\bar{Z}_i}{r_{31}u_i + r_{32}v_i + r_{33} + t_z/\bar{Z}_i} \\ \frac{r_{21}u_i + r_{22}v_i + r_{23} + t_y/\bar{Z}_i}{r_{31}u_i + r_{32}v_i + r_{33} + t_z/\bar{Z}_i} \end{bmatrix}, \quad (3.19)$$

\mathbf{w}_{ik} is EPF for frame k , and $\bar{\mathbf{X}}_{nk}$ is the average motion vector of the feature points within group n . The derivative matrix, $\frac{\partial \mathbf{f}_i}{\partial \mathbf{X}_{ik}}$, converts the motion uncertainty into feature uncertainty in the image plane. Figure 3-7 illustrates prediction of uncertainty field of the surface feature points (projection field) with EPF, which generates the searching areas for candidate correspondence in the current frame.

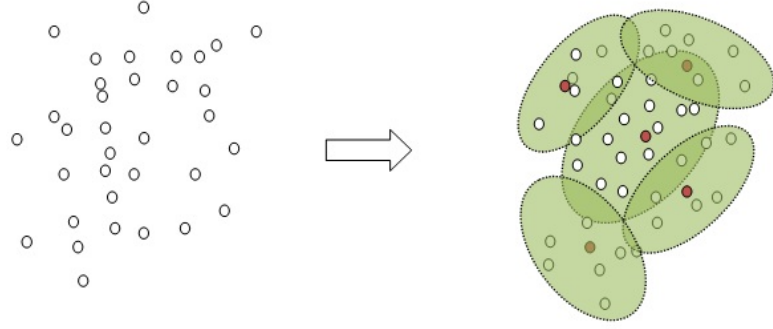


Figure 3-7. Motion field prediction with uncertainty error. The red points represent the motion centers of the different feature groups. The green regions represent the prediction regions with uncertainty error.

By defining the measurement at frame $k-1$ as $\mathbf{p}'_{i(k-1)} = (u_{(k-1)}, v_{(k-1)}, Z_{(k-1)})$, Equation 3.16 can be written as

$$f(P_k, P_{k-1}, Z_{k-1}, \mathbf{R}_{k-1}, \mathbf{T}_{k-1}) = f(P_k, P'_{k-1}, Z_{k-1}, \mathbf{X}_{n(k-1)}). \quad (3.20)$$

The feature point p_i at frame k can then be updated by estimating motion vector \mathbf{X}_k through Equation 3.21 and 3.22

$$\begin{aligned} f(P_k, P'_{k-1}, Z_{k-1}, \mathbf{X}_{n(k-1)}) + \frac{\partial f_i}{\partial \mathbf{X}_{nk}} \Big|_{\mathbf{X}_{nk}=\bar{\mathbf{X}}_{nk}} (\mathbf{X}_{nk} - \bar{\mathbf{X}}_{nk}) + \\ \frac{\partial f_i}{\partial \mathbf{p}'_{ik}} \Big|_{\mathbf{p}_{ik}=\bar{\mathbf{p}}'_k} (\mathbf{p}'_{ik} - \bar{\mathbf{p}}'_k) \approx 0, \end{aligned} \quad (3.21)$$

$$\begin{aligned}
 & -f(P_k, P'_{k-1}, Z_{k-1}, \mathbf{X}_{n(k-1)}) + \frac{\partial f_i}{\partial \mathbf{X}_{nk}}|_{\mathbf{x}_{nk}=\bar{\mathbf{x}}_{nk}} \bar{\mathbf{X}}_{nk} \\
 & \approx \frac{\partial f_i}{\partial \mathbf{X}_{nk}}|_{\mathbf{x}_{nk}=\bar{\mathbf{x}}_{nk}} \mathbf{X}_{nk} + \frac{\partial f_i}{\partial \mathbf{p}'_{ik}}|_{\mathbf{p}_{ik}=\bar{\mathbf{p}}'_k} (\mathbf{p}'_{ik} - \bar{\mathbf{p}}'_k),
 \end{aligned} \tag{3.22}$$

which can be written as

$$\hat{Z}_{ik} = \mathbf{H}_i \mathbf{X}_{nk} + \mathbf{v}_{ik}, \tag{3.23}$$

where

$$\hat{Z}_{ik} = -f(P_k, P'_{k-1}, Z_{k-1}, \mathbf{X}_{n(k-1)}) + \frac{\partial f_i}{\partial \mathbf{X}_{nk}}|_{\mathbf{x}_{nk}=\bar{\mathbf{x}}_{nk}} \bar{\mathbf{X}}_{nk},$$

$$\mathbf{H}_i = \frac{\partial f_i}{\partial \mathbf{X}_{nk}}|_{\mathbf{x}_{nk}=\bar{\mathbf{x}}_{nk}}, \tag{3.24}$$

$$\mathbf{v}_{ik} = \frac{\partial f_i}{\partial \mathbf{p}'_{ik}}|_{\mathbf{p}'_{ik}=\bar{\mathbf{p}}'_k}. \tag{3.25}$$

Covariance matrix \mathbf{G}_i of the observation noise sequence \mathbf{v}_{ik} is given by

$$\begin{aligned}
 \mathbf{G}_i &= E(\mathbf{p}_{ik} - \bar{\mathbf{p}}_{ik})(\mathbf{p}_{ik} - \bar{\mathbf{p}}_{ik})^T \\
 &= \frac{\partial f_i}{\partial \mathbf{p}'_{ik}}|_{\mathbf{x}_k=\bar{\mathbf{x}}_{nk}} * \mathbf{w}'_i * \left[\frac{\partial f_i}{\partial \mathbf{p}'_{ik}}|_{\mathbf{p}_k=\bar{\mathbf{p}}_{nk}} \right],
 \end{aligned} \tag{3.26}$$

where

$$\frac{\partial f_i}{\partial \mathbf{p}'_{ik}} = \begin{bmatrix} \frac{\partial f_1}{\partial u_i} & \frac{\partial f_1}{\partial v_i} & \frac{\partial f_1}{\partial z_i} \\ \frac{\partial f_2}{\partial u_i} & \frac{\partial f_2}{\partial v_i} & \frac{\partial f_2}{\partial z_i} \end{bmatrix}, \tag{3.27}$$

and

$$\mathbf{w}' = \begin{bmatrix} \mathbf{w}_i & 0 \\ 0 & \sigma_{zik}^2 \end{bmatrix}. \tag{3.28}$$

Thus, according to definition of Kalman filter, the Kalman gain can be derived as,

$$\mathbf{K}_{ik} = \mathbf{w}_{ik}(\mathbf{H}_{ik})^T [\mathbf{G}_{ik} + \mathbf{H}_{ik} \mathbf{w}_{ik}(\mathbf{H}_{ik})^T]^{-1}. \tag{3.29}$$

Estimate of the motion vector and covariance matrix can be calculated by

$$\mathbf{X}_k = \mathbf{X}_k^- - \mathbf{K}_{ik} \mathbf{H}_{ik}, \quad (3.30)$$

and

$$\mathbf{w}_{ik} = (1 - \mathbf{K}_{ik} \mathbf{H}_{ik}) \mathbf{w}_{ik}^-, \quad (3.31)$$

respectively, where \mathbf{X}_k^- is the forecast motion vector, \mathbf{w}_{ik}^- is the forecast error covariance.

3.3.3 Feature Point Matching Based on Minimal Bending Energy

In Section 3.2, the feature points in each frame are estimated by predicting their motion vectors and related projection field with prediction uncertainty. With the Kalman gain updated in EKF tracking, the motion vectors are updated from their previous states (Equation 3.29). However, the motion vectors are also affected by local surface properties due to non-rigid motion of human skin. In this section, a bending energy model is used to optimally determine the optimal corresponding feature points from the updated feature candidates in the searching windows.

Under an assumption that the patient (skin) surface deforms as little as possible between the two sequential captured frames (within small time intervals), the candidate feature points which have the minimum bending energy with their corresponding feature points in the previous frame are considered as the expected correspondences in the current frame. This can be modeled by

$$\hat{p}_2 \stackrel{argmin}{=} [\epsilon(p_1) - \epsilon(p_{2i})], \quad p_{2i} \in W, \quad (3.32)$$

where $\epsilon(p_1)$ and $\epsilon(p_{2i})$ are bending energy at the feature points in the previous and current frame respectively. \hat{p}_2 is the estimated correspondence of the feature points in the current frame. W is a user-defined searching region which contains candidates of the updated feature points from EKF tracking.

The motion centers of CFC that are acquired from multi-EKF tracking procedure (Section 3.3.2) are considered as the initial input for calculation of the centers of searching windows. Match-weighted measure (Shi et al., 2000) can be utilized to examine the goodness and uniqueness of feature points matching.

The local surface patches of patient skin are modeled as thin flexible plates. The potential energy of an ideal thin flexible plate (elastic material) is a measure of the strain energy of deformation from its equilibrium state. It is a function of the two principal curvatures of the plate, and is invariant to 3D rotation and translation:

$$\epsilon = A\left(\frac{\kappa_1^2 + \kappa_2^2}{2}\right), \quad (3.33)$$

where A is a material related constant for the particular plate. κ_1 and κ_2 are two principal curvatures for the surface patch of interest:

$$\kappa_1 = H + \sqrt{H^2 - K}, \quad (3.34)$$

$$\kappa_2 = H - \sqrt{H^2 - K}. \quad (3.35)$$

e, g, f, E, G and F are curvature parameters defined in the literature (Ait-Aoudia et al., 2006) for computation of H and K ,

$$K = \det|\beta| = \frac{eg - f^2}{EG - F^2}, \quad (3.36)$$

$$H = \frac{\text{tr}[\beta]}{2} = \frac{eg - 2fF + gE}{2(EG - F^2)}, \quad (3.37)$$

where β is the Weingarten Mapping Matrix,

$$|\beta| = - \begin{bmatrix} e & f \\ f & g \end{bmatrix} \begin{bmatrix} E & F \\ F & G \end{bmatrix}^{-1}. \quad (3.38)$$

Assume κ_1 and κ_2 are two principal curvatures at surface point p_1 . The principal curvatures of the deformed surface patch at their corresponding point p_2 are κ'_1 and κ'_2 . The bending energy ϵ_t can then be used as a matching criterion to estimate the point correspondence. According to bending energy Equation 3.33, a surface potential energy is generated when bending a curved surface patch to a newly deformed state as:

$$\epsilon_t = A \left(\frac{(\kappa_1 - \kappa'_1)^2 - (\kappa_2 - \kappa'_2)^2 + (\kappa_1 \kappa_2 - \kappa'_1 \kappa'_2)}{2} \right), \quad (3.39)$$

Under the assumption that the surface patch around the each feature point deforms slightly and locally in the sequential frames, a physically-plausible search region is constructed on the updated surface S_2 at the frame $(k+1)$ at each feature candidate. The updated feature points that are acquired from Kalman filter tracking are taken as initial input for building searching regions W . W consists of the points that have connectivity distance less than a threshold M at the frame $(k+1)$.

Measure of bending energy between the surface patch around point p_{ik} and surface patches surrounding points $p_{i(k+1)}$ are computed. The point $\hat{p}_{i(k+1)}$ that has the minimum bending energy within the searching window W is chosen as the point correspondence to point p_{ik} . By Equation 3.34 to 3.37, the shape properties are calculated in terms of the principal curvatures κ_1 and κ_2 . Point searching is performed within the plausible region W on the updated surface S_2 . The point

\hat{p}_2 in the frame $(k+1)$, which has the closest local shape properties to point p_1 in the frame k , is selected as the optimal correspondence of p_1 (Equation 3.4).

$$d(\hat{p}_2) \stackrel{argmin}{=} \frac{1}{2} |[\kappa_1(p_1) - \kappa'_1(p_2)]^2 - [\kappa_2(p_1) - \kappa'_2(p_2)]^2 + [\kappa_1(p_1)\kappa_2(p_1) - \kappa'_1(p_2)\kappa'_2(p_2)]|. \quad (3.40)$$

3.4 Summary

Our contributions to ProCam calibration are elucidated in Chapter 3. The first topic (Section 3.1 and 3.2) is camera and projector calibration used for ProCam system calibration and projection calibration. An improved structured light pattern with phase shift is developed for noise suppression and accuracy enhancement. The second topic (Section 3.3) is on dynamical calibration of the ProCam system. A hybrid algorithm combining with EKF and minimal energy estimation is proposed. EKF provides the updated ProCam pixel correspondence between the object surface and camera image, while the minimal energy refines the corresponding points on the camera image to find the optimal point correspondence. Graphics Processing Unit (GPU) based parallel computing can be used to speed up the computationally expensive process. The implementation details and experiment results of using a Compute Unified Device Architecture(CUDA) based GPU parallel computing are described in Section 7.4. This is probably the first method that can achieve real-time ProCam calibration without pre-generating embedded structured patterns in the source image sequence.

Chapter 4

GEOMETRIC AND RADIOMETRIC CORRECTION

In projector-based AR display, the projection image is usually distorted by an arbitrary surface. The surfaces' irregular 3D geometry and textures are the main factors to cause distortion including the geometric and radiometric distortion (Figure 4-1). In order to create an immersive AR environment where the computer-generated imagery appears correctly and seamlessly in its surrounding real environment, geometric and radiometric correction are usually required. The image correction is particularly challenging for a projector-based direct AR display (Table 1.1). This is because its AR effect depends on synthetic interaction between the projection and the actual object surface. Projector-based AR display requires a direct overlay of the virtual scene (projection imagery) onto the real object surface rather than simply merging their images on a computer screen.

Geometric alignment between the projection image and the surface is traditionally achieved by physical construction, where projectors are mounted on the

sophisticated controllable mounts (Bimber and Raskar, 2005). Even with these projector mounts, subsequent manual adjustment is needed to refine the projector’s position and orientation to ensure an accurate alignment of the projected imagery. A high-quality projection display requires the projection without sacrificing the photometric quality on any sophisticated background. The fundamental problem of radiometric correction on an arbitrary surface is that the surface cannot be assumed to be white and matte.

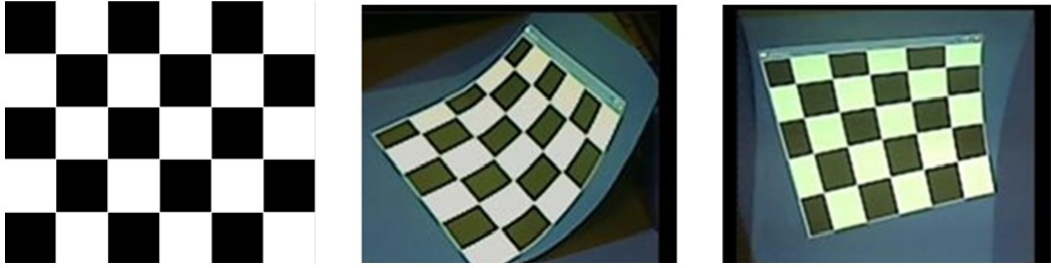


Figure 4-1. Geometric and radiometric distortion.

In this chapter, an image pre-warping method is developed instead of physically adjusting the projector’s position and orientation for projection geometric correction. Radiometric correction can be achieved by generating compensation images to adapt to surface’s irregular texture, reflectance properties and environmental lighting. The principles of viewer-dependent geometric and radiometric correction are presented in section 4.1 and 4.2. The proposed methods using piecewise texture mapping for both geometric and radiometric correction are then described in section 4.3.

4.1 Geometric Correction

Current geometric correction methods for direct SAR could be classified by the type of the screen surfaces: parametric and nonparametric surfaces. The parametric surfaces have known geometric parameters of the display surfaces. The spatial

relationship between the projector image and the projection image can thus be easily derived from the known parameters (Section 2.3). A pre-warping correction model for parametric surfaces is proposed by Majumder and Brown (Majumder and Brown, 2007), which establishes pixel mapping between the projector image and projection image. For nonparametric surfaces, the above parameterization is not known in advance. Since most of the display surfaces have unknown geometric parameters especially for the arbitrary display surfaces, the pre-warping methods for the nonparametric surfaces has a wide range in practice.

Although the warping technique for projection image correction is used in current projectors, it is mostly used for keystone correction. For example, most of current projectors may have built-in function of keystone correction which can warp the input image to mitigate distortions caused by off-axis projection. This warping correction for keystone is typically applied along the horizontal scanline only with an assumption that the projector is placed on a table and keystone only affects the horizontal scanlines.

The proposed method in this section involves a pre-warping concept and efficient pre-warping algorithm that enables the geometric alignment of the projection images to correctly adapt to a non-planar surface. With this method, the sequential structured-light projection is not required and the feature points can be flexibly assigned by users. In addition, the projector and camera do not have to be set coaxially.

4.1.1 Principle of Viewer-dependent Pre-warping

Projection's geometric distortion is caused by projective distortion and surface irregular topography. If the display surface is planar and projection light is perpendicular to the surface, the projection on the screen surface may appear hardly

distorted to the observers in front of the screen. When a projector image is projected on a display surface that is not perpendicular to the projector axis, the projective distortion and the lens distortion may corrupt the projection image. When a projector image is projected onto a display surface with irregular unevenness, surface topography is the main factor to cause projection distortion.

For a planar projection, homography is used to describe the transformation from the source image to the projected image in the world coordinate, from the projected image to the captured image in the camera, as Equation 4.1 and 4.2 (Wen et al., 2010a):

$$\begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix} = H_p \begin{bmatrix} w_x \\ w_y \\ 1 \end{bmatrix}, \quad (4.1)$$

$$\begin{bmatrix} c_x \\ c_y \\ 1 \end{bmatrix} = H_c \begin{bmatrix} w_x \\ w_y \\ 1 \end{bmatrix}. \quad (4.2)$$

$(w_x, w_y, 1)^T$ is an image point projected on a planar surface. $(p_x, p_y, 1)^T$ and $(c_x, c_y, 1)^T$ are the corresponding points located in the projector's and camera's respectively. From Equation 4.1 and 4.2, a spatial relationship from the source image to the captured image can be established in Equation 4.3 (Figure 4-2):

$$\begin{bmatrix} p_x \\ p_y \\ 1 \end{bmatrix} = H_p H_c^{-1} \begin{bmatrix} c_x \\ c_y \\ 1 \end{bmatrix}. \quad (4.3)$$

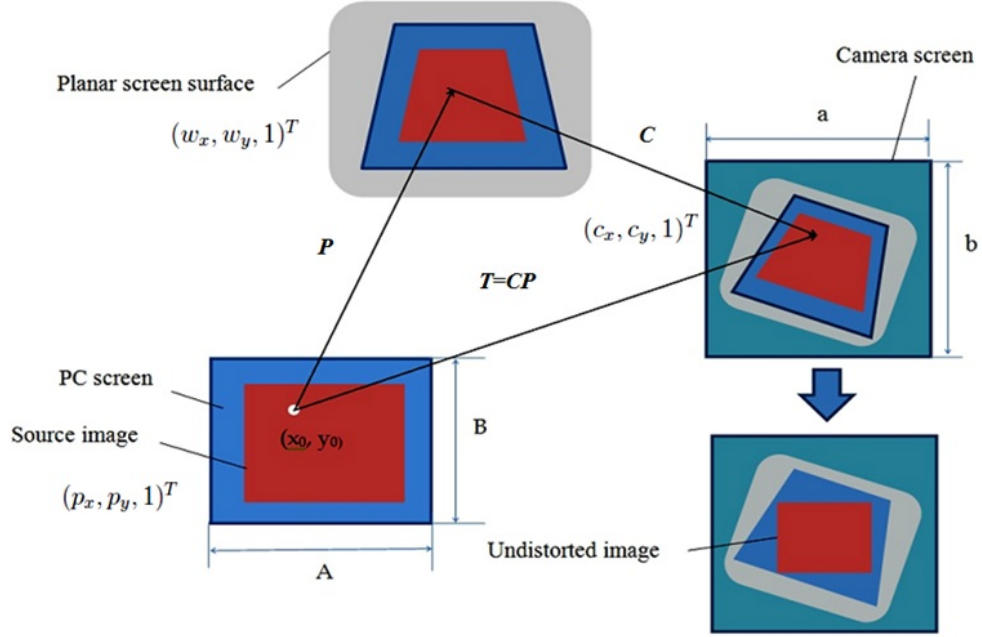


Figure 4-2. Viewer-dependent geometric correction.

The geometric correction process is illustrated in Figure 4-2. There are two types of correction models: viewer dependent and viewer independent. For viewer-dependent correction, the source image should be pre-warped as $T^{-1}S$, where S is the source image and T is the mapping function between the source image and the captured image. For viewer-independent correction, the source image should be warped as $P^{-1}S$, where P is the mapping function between the source image and the projected image. Equation 4.4 and 4.5 illustrate these correction processes.

$$CP(T^{-1}S) = CP(P^{-1}C^{-1})S = S, \quad (4.4)$$

$$P(P^{-1}S) = P(CT^{-1})S = S. \quad (4.5)$$

4.1.2 Piecewise Pre-warping

Although Equation 4.3 is not directly applicable when the screen is non-flat, the function $H_p H_c^{-1}$ can be found easily. An idea of replacing $H_p H_c^{-1}$ by piecewise linear mapping function is proposed to relate the non-flat object surface, camera

and projector. With the piecewise linear method, the homography for source image can be computed piecewise with the defined feature points. The source image can then be pre-warped piece by piece. This enables correction of the projected image on a non-planar surface by connecting all these corrected piecewise regions. A quadrilateral piece can be used as the unit region to compute the homography between the source image and projected image, the captured image and source image. The region connection could be compensated by bilinear interpolation.

The point-correspondence between the image pairs can be acquired from the ProCam system calibration results. For every piece region, the corresponding feature points are established with the respective relationship defined by Equation 4.6, 4.7 and 4.8. i is the i^{th} piecewise region in the source image, projected image and the captured image. The other parameters have the same meaning as those in Equation 4.1, 4.2 and 4.3.

$$\begin{bmatrix} p_{ix} \\ p_{iy} \\ 1 \end{bmatrix} = H_{pi} \begin{bmatrix} w_{ix} \\ w_{iy} \\ 1 \end{bmatrix}, \quad (4.6)$$

$$\begin{bmatrix} c_{ix} \\ c_{iy} \\ 1 \end{bmatrix} = H_{ci} \begin{bmatrix} w_{ix} \\ w_{iy} \\ 1 \end{bmatrix}, \quad (4.7)$$

$$\begin{bmatrix} p_{ix} \\ p_{iy} \\ 1 \end{bmatrix} = H_{pi} H_{ci}^{-1} \begin{bmatrix} c_{ix} \\ c_{iy} \\ 1 \end{bmatrix}. \quad (4.8)$$

Given the correspondence between (p_{ix}, p_{iy}) and (c_{ix}, c_{iy}) , homography for a

certain piecewise region can be computed by setting up a system of linear equations and solving the unknown parameters with the least-squares method.

Assume the homography for piecewise region R_i is $H_i (H_{pi}, H_{ci}, H_{pi}H_{ci}^{-1})$ which relates the two different space domains (S_x, S_y) and (D_x, D_y) . The set of correspondence between these domain is denoted by $(S_{ix}, S_{iy}) \rightarrow (D_{ix}, D_{iy})$. The (S_{ix}, S_{iy}) and (D_{ix}, D_{iy}) are related by a homography defined by nine parameters $(h_1, h_2, \dots, h_8, h_9)$,

$$\begin{bmatrix} D_{ix} \\ D_{iy} \\ 1 \end{bmatrix} = \begin{bmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{bmatrix} \begin{bmatrix} S_{ix} \\ S_{iy} \\ 1 \end{bmatrix}, \quad (4.9)$$

then we have,

$$D_{ix} = S_{ix}h_1 + S_{iy}h_2 + h_3, \quad (4.10)$$

$$D_{iy} = S_{ix}h_4 + S_{iy}h_5 + h_6, \quad (4.11)$$

$$1 = S_{ix}h_7 + S_{iy}h_8 + h_9. \quad (4.12)$$

Because the homography is defined with an arbitrary scaling, h_9 can be set to 1. Then the following two equations can be derived from Equation 4.10, 4.11 and 4.12 with respect to the eight unknowns h_1, \dots, h_8 :

$$S_{ix}h_1 + S_{iy}h_2 + h_3 - D_{ix}S_{ix}h_7 - D_{ix}S_{iy}h_8 - D_{ix} = 0, \quad (4.13)$$

$$S_{ix}h_4 + S_{iy}h_5 + h_6 - D_{iy}S_{ix}h_7 - D_{iy}S_{iy}h_8 - D_{iy} = 0. \quad (4.14)$$

Given a set of N corresponding points $(S_{ix}, S_{iy}) \rightarrow (D_{ix}, D_{iy})$ in piecewise region R_i , the following linear system can be set up from Equations 4.13 and 4.14:

$$\mathbf{A}\mathbf{h}_i = \mathbf{b}, \quad (4.15)$$

where

$$\mathbf{A} = \begin{bmatrix} S_{ix1} & S_{iy1} & 1 & 0 & 0 & 0 & -D_{ix1}S_{ix1} & -D_{ix1}S_{iy1} \\ 0 & 0 & 0 & S_{ix1} & S_{iy1} & 1 & -D_{iy1}S_{ix1} & -D_{iy1}S_{iy1} \\ S_{ix2} & S_{iy2} & 1 & 0 & 0 & 0 & -D_{ix2}S_{ix2} & -D_{ix2}S_{iy2} \\ 0 & 0 & 0 & S_{ix2} & S_{iy2} & 1 & -D_{iy2}S_{ix2} & -D_{iy2}S_{iy2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S_{ixN} & S_{iyN} & 1 & 0 & 0 & 0 & -D_{ixN}S_{ixN} & -D_{ixN}S_{iyN} \\ 0 & 0 & 0 & S_{ixN} & S_{iyN} & 1 & -D_{iyN}S_{ixN} & -D_{iyN}S_{iyN} \end{bmatrix}, \quad (4.16)$$

$$\mathbf{h}_i = [h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8]^T, \quad (4.17)$$

$$\mathbf{b} = [D_{ix1}, D_{iy1}, D_{ix2}, D_{iy2}, \dots, D_{ixN}, D_{iyN}]^T. \quad (4.18)$$

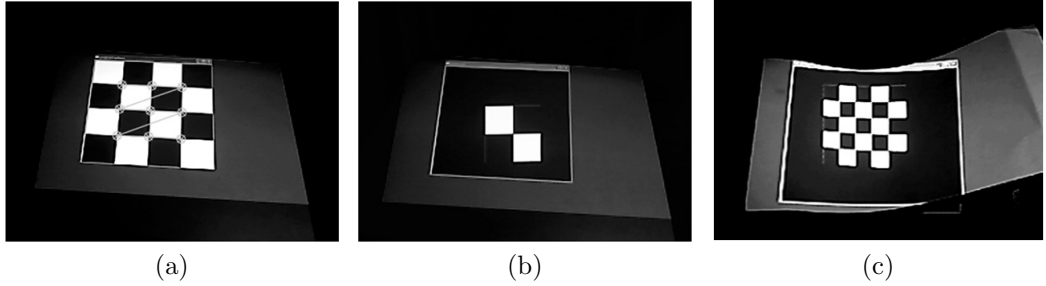


Figure 4-3. Geometric correction on a planar surface (b) and curved surface (c) with piecewise pre-warping method. The piecewise regions are defined by the four feature points in quadrilaterals (a) in this example.

In order to find the optimal vector \mathbf{h}_i , its parameters are computed by minimizing the Euclidean norm $\|Ah_i b\|_2$. When more than four projector-camera correspondences within a piecewise region are used, the matrix \mathbf{A} has eight columns which have the same number as the unknown parameters in \mathbf{A} . The number of rows is two times the number of correspondences N . Thus, \mathbf{A} is a $2N \times 8$ matrix. To solve \mathbf{h}_i , singular value decomposition (SVD) (Majumder and Brown, 2007) can be used to compute $(ATA)x = A^T h$ using the pseudoinverse $A^T A$ such that

$\mathbf{h}_i = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$. An example of piecewise geometric correction for chessboard correction on a planar and curved surface is shown as Figure 4-3.

4.2 Radiometric Correction

In an actual environment, the color of projection display is easily affected by variation of the device and environment factors. The variation includes intra-projector color variation, environment lighting, curved display surface, surface texture and inter-reflection. Radiometric correction involves a sophisticated color model construction. In this section, a radiometric correction model is proposed, based on the Majumder's lemmas (Majumder and Brown, 2007) and Nayar's color model (Nayar et al., 2003), to correct photometric distortion in surgical AR environment (adapting to patient skin surface). This color model establishes color transformation among the projector, camera and environment lighting. Based on this ProCam radiometric model, a radiometric compensation method is developed to approximate projection display to the color in the source projector image considering surface reflectance and environment lighting.

4.2.1 Radiometric Model for ProCam

According to the Majumder and Brown's lemmas (Majumder and Brown, 2007), intensity of the color light has the following properties:

Lemma 4.1. The intensities of the colors that lie on a ray originating at the origin of the XYZ space have the same chromaticity coordinates, and their intensities are related by a scale factor.

Lemma 4.2. Addition of two colors results in a color whose intensity is the addition of the intensities of the comprising colors.

Lemma 4.3. When two colors are added, the chromaticity coordinates of the constituent colors are added in proportion to their intensities to create the chromaticity coordinate of the new color.

Thus, irradiance of the projection display in the camera direction can be represented by

$$S(\lambda)_{R,G,B} = s(\lambda)[P(\lambda)_{R,G,B} + E(\lambda)_{R,G,B}], \quad (4.19)$$

where $P(\lambda)_{R,G,B}$ is the projector light with R, G, B channel, and $E(\lambda)_{R,G,B}$ is the corresponding environment light reflected by the surface in the camera direction. $s(\lambda)$ is the spectral reflectance of the irradiated surface. (The spectrum s is a function of wavelength λ , $s(\lambda)$.) According to the radiometric response model within a projector (Nayar et al., 2003), the transformation from the pixel value I_s in the source image to the projection light $P(\lambda)_{R,G,B}$ can be described by

$$P(\lambda)_{R,G,B}(\lambda) = e(I_p)_{R,G,B} f_p(\lambda). \quad (4.20)$$

where the $e(I_p)_{R,G,B}$ is the radiometric response of the electronics of the projector, I_p is the brightness of the projector image converted from computer input I_s , $I_p = N_{R,G,B}(I_s)$, and $f_p(\lambda)$ is the spectral response of the projector channel.

With the input of the projection display on the surface, its irradiance can be converted to the pixel value of the camera image by radiometric response of the electronics of the camera $f_c(\lambda)$ and frame-grabber $G_{R,G,B}$,

$$I_o = G_{R,G,B}(C_{R,G,B} f_c(\lambda)), \quad (4.21)$$

where the quantum efficiency $q_{R,G,B}(\lambda)$ is involved in computation of the captured light $C_{R,G,B}$,

$$C_{R,G,B} = I_p \int f_p(\lambda) s(\lambda) q_{R,G,B}(\lambda) d(\lambda). \quad (4.22)$$

The above equations describe the brightness conversion within the ProCam system, among computer, projector and camera. Based on this relationship, a radiometric correction model is proposed to generate appropriate pixel values in R, G, B channel to compensate the radiometric distortion on the projection surface.

$$C_{R,G,B} = I_p \int f_p(\lambda) s(\lambda) q_{R,G,B}(\lambda) d(\lambda). \quad (4.23)$$

Image $C = [C_R \ C_G \ C_B]^T$ acquired by camera from the projected image $P = [P_R \ P_G \ P_B]^T$ is affected by the surface reflectance and environment lighting. Assume projector and camera each has three color channels (R, G, B). Equation 4.24 and 4.25 illustrate the radiometric transformation from the projector to camera via the screen surface based on the above photometric models:

$$\mathbf{C} = \mathbf{F}(\mathbf{E} + \mathbf{M}\mathbf{P}), \quad (4.24)$$

$$\begin{bmatrix} C_R \\ C_G \\ C_B \end{bmatrix} = \begin{bmatrix} F_R E_R \\ F_G E_G \\ F_B E_B \end{bmatrix} + \begin{bmatrix} F_R M_{RR} & M_{RG} & M_{RB} \\ M_{GR} & F_G M_{GG} & M_{GB} \\ M_{BR} & M_{BG} & F_B M_{BB} \end{bmatrix} \begin{bmatrix} P_R \\ P_G \\ P_B \end{bmatrix}. \quad (4.25)$$

where the vector \mathbf{F} is the reflectance of surface which is assumed to be constant for each channel,

$$\mathbf{F} = \begin{bmatrix} F_R & 0 & 0 \\ 0 & F_G & 0 \\ 0 & 0 & F_B \end{bmatrix}, \quad (4.26)$$

\mathbf{E} is the contribution of the environment lighting to the captured image. The color mixing matrix \mathbf{M} describes the radiometric interaction between the projector and the camera. The color mixing matrix \mathbf{M} can be derived by projecting two single-channel images with channel values of 255 and 0 respectively for monotonic radiometric calibration. To compute \mathbf{M} for the color case, normalization of the elements in \mathbf{M} should be conducted first to eliminate the influence of the surface reflectance. The diagonal elements in \mathbf{M} are evaluated to unity, $M_{R,G,B} = 1$. By comparing images that are different in one of the three channels, the other elements of M_{ij} can be acquired for the corresponding channels (R, G, B). Based on the piecewise linear method described in Section 4.1, the matrix \mathbf{F} , \mathbf{E} and \mathbf{M} can be replaced by \mathbf{F}_i , \mathbf{E}_i and \mathbf{M}_i as Equation 4.6, 4.7 and 4.8.

4.2.2 Radiometric Compensation

In the radiometric correction model described in Equation 4.25, the elements in the matrix of reflectance of the surface irradiance \mathbf{F} , environment lighting \mathbf{E} and color mixing matrix \mathbf{M} are difficult to be measured separately. However, the combination of \mathbf{FE} and \mathbf{FM} can be computed if the corresponding pairs of projector image color \mathbf{P} and camera image color \mathbf{C} are used. In this radiometric correction, the projector image is compensated with the difference between the source image and projection image color. The camera image provides the color feedback of the projection display.

We assume that the environment light is consistent and ambient. When the projector light is turned off, we can assume the projector input is set to zero, Equation 4.24 and 4.25 can be rewritten as:

$$\mathbf{C} = \mathbf{FE}, \quad (4.27)$$

$$\begin{bmatrix} C_R \\ C_G \\ C_B \end{bmatrix} = \begin{bmatrix} F_R E_R \\ F_G E_G \\ F_B E_B \end{bmatrix}. \quad (4.28)$$

Thus, the matrix composite \mathbf{FE} can be directly acquired from the camera image color \mathbf{C} .

In comparison with the brightness of environment lighting, projector is considerably brighter when performing the projection on a textured and colored surface. When the environment lighting is set to zero (e.g. the room lighting is turned off), the matrix pair \mathbf{FE} can be considered as zero under the projection lighting environment. In this case, Equation 4.24 and 4.25 can be rewritten as:

$$\begin{bmatrix} C_R \\ C_G \\ C_B \end{bmatrix} = \begin{bmatrix} F_R M_{RR} & M_{RG} & M_{RB} \\ M_{GR} & F_G M_{GG} & M_{GB} \\ M_{BR} & M_{BG} & F_B M_{BB} \end{bmatrix} \begin{bmatrix} P_R \\ P_G \\ P_B \end{bmatrix}. \quad (4.29)$$

To compute elements of the matrix combination \mathbf{FM} , two images with pixel color only different in one of the three channels (R, G, B) are used to compute their differences in the other two channels. Suppose that the two images with pixel color differing in red channel. The channel value in image I_1 and I_2 are denoted by

$$\mathbf{C}_{1(R_1, G_1, B_1)} = \mathbf{FMP}_{1(R_1, G_1, B_1)}, \quad (4.30)$$

and

$$\mathbf{C}_{2(R_1, G_1, B_1)} = \mathbf{FMP}_{2(R_2, G_1, B_1)}. \quad (4.31)$$

Since pixel values are only different in channel R between the two images, thus we have

$$M_{RR} = \Delta C_R / \Delta P_R, \quad (4.32)$$

$$M_{GR} = \Delta C_G / \Delta P_R, \quad (4.33)$$

$$M_{BR} = \Delta C_B / \Delta P_R. \quad (4.34)$$

In this way, we can compute the elements of matrix \mathbf{M} by projecting every two images with one channel different. The diagonal elements are set equally to unity ($M_{RR} = 1$) to facilitate calculation of the other elements in the row. Once the matrix combination $\mathbf{F}\mathbf{E}$ and $\mathbf{F}\mathbf{M}$ are acquired, the expected projection image color can be generated by

$$\mathbf{P}_{exp} = (\mathbf{F}\mathbf{M})^{-1}(\mathbf{C} - \mathbf{F}\mathbf{E}). \quad (4.35)$$

An example of radiometric correction is shown in Figure 4-4. The region of interest (ROI) for radiometric compensation is the "liver". The expected color display in ROI can be changed according to pixel values in the source image or redefined by users.

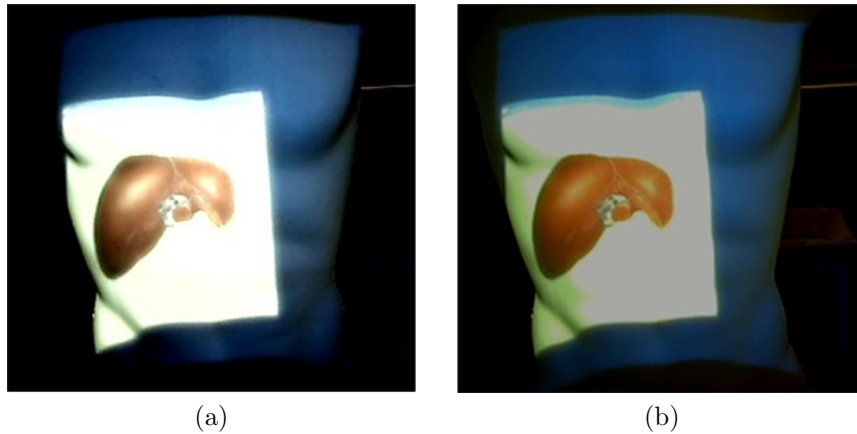


Figure 4-4. Radiometric correction of a liver model on a mannequin body: (a) before correction, (b) after correction.

4.3 Texture Mapping for Pixel Value Correction

The piecewise pre-warping method in Section 4.1 and 4.2 for geometric and radiometric correction uses the parametric functions. Although this method can be directly implemented based on the correction principles (geometric and radiometric correction model), estimation of the parametric elements in the matrices usually requires massive data fitting computation (Equation 4.16) to achieve a satisfying correction accuracy. Its implementation may also be a time-consuming procedure. In this section, a texture mapping method is developed, based on the previous correction principles (Section 4.1), to accurately map the pixel value from the camera coordinates to projector coordinates. The texture mapping method has advantages of fast calibration and high accuracy. The recent development on GPU-based programming (Olmedo et al., 2012) enables this method more efficient for the practical application. In addition, this texture mapping method can use the results of previous system calibration (structured-light calibration).

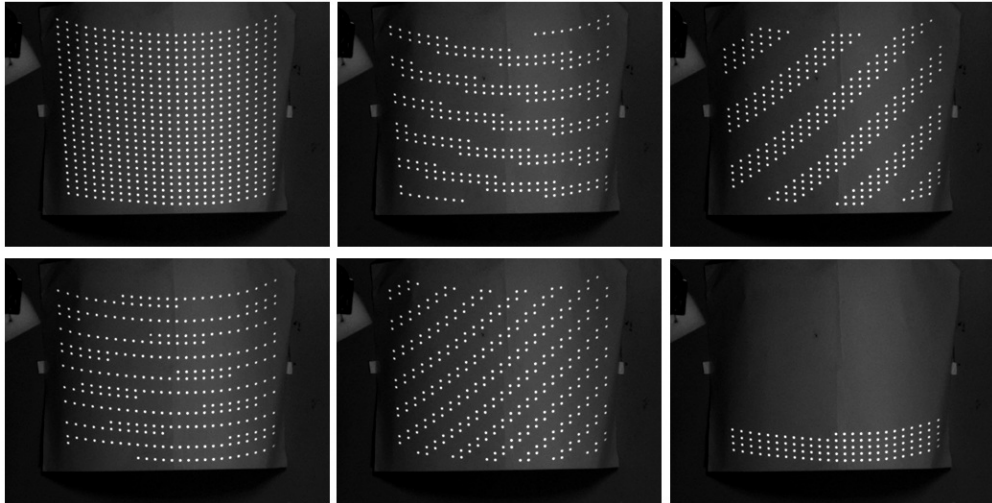


Figure 4-5. Blob cluster are projected to establish the texture mapping.

To acquire the pixel mapping between the projector and camera coordinates, the system calibration method in Chapter 3 can provide a satisfying data mapping

(lookup table). In this study, a serial of blob-matrix images is developed to generate the data mapping (Figure 4-5). This can be achieved by projecting equally spaced features and detecting their positions in the observed camera image. The size of each Gaussian blob can be defined by users (Majumder and Brown, 2007).

In order to determine the correspondence between the projected and camera image, each Gaussian blob is assigned with an ID in bits representing its position in the cluster of blobs. The texture mapping of the projector pixel to the camera pixel is based on the tessellation of the projection image. Assume the projector and camera image coordinates are (x, y) and (u, v) respectively. The number of their pixel correspondence is n and denoted by $f: (u_k, v_k) \leftrightarrow (x_k, y_k), 0 \leq k \leq n$. The (u_k, v_k) are tessellated to create triangulation in the camera coordinates. The correspondences of (u_k, v_k) in the projector image (x_k, y_k) are connected with triangulation in the same way as (u_k, v_k) to generate the corresponding triangle regions in the projector image. This texture mapping method uses the triangulation in the camera image to find the corresponding blobs and assign the pixel values to generate the pre-warped projector image.

In order to find pre-warped triangulation in the projector image, the blobs' coordinates in the camera image are normalized first within a user-defined rectangle region. In this region, the expected projection display are planned and tessellated by triangulation of the blobs' coordinates. With triangulation and texture mapping of the pixel values of the camera image, the corresponding triangulation in the projector image can be found.

This texture mapping method allows projection correction on a non-parametric surface with nonlinear warping of the source image (Figure 4-6). This method can easily achieve subpixel accuracy even with lens distortion in the projector. It relies on direct texture mapping with pixel value correction, and thus avoids errors

resulting from the data fitting process in Section 4.1.

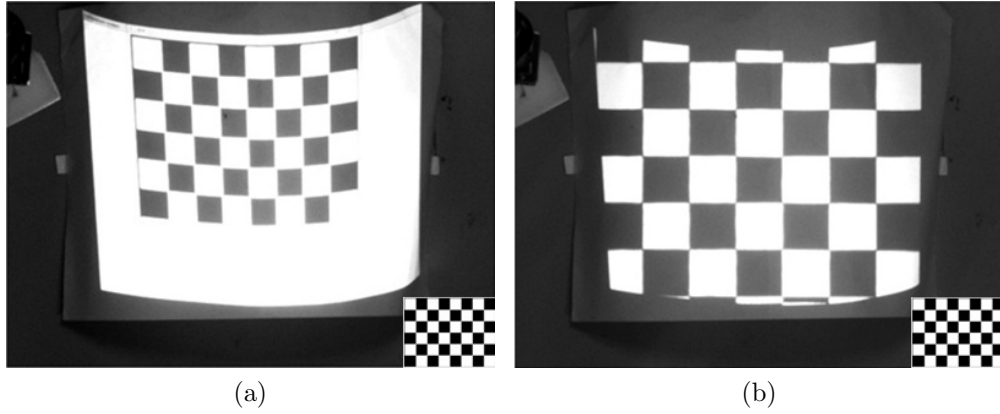


Figure 4-6. Projection correction on a curved surface based on texture mapping: (a) projection distortion (checkerboard pattern) caused by the curved surface; (b) projection on the curved surface.

4.4 Summary

Chapter 4 reports our progress on geometric and radiometric correction for the construction of a ProCam-based AR environment on an arbitrary surface. The principle of image pre-warping and the piecewise linear correction method based on the principle are firstly described for geometric correction in Section 4.1. In this method, the homography for the corresponding piecewise regions is assumed to be consistent. With the pre-warping principle and piecewise method, a radiometric correction model and radiometric compensation method are developed for projection correction on a textured surface. The relationship among environmental lighting, form factor, screen color and surface textures is considered in the proposal radiometric correction model. In the last section, the piecewise linear method is improved by a texture mapping method with pixel value correction, which enhances the accuracy of the projection correction.

Chapter 5

REGISTRATION

In the construction of an AR system, registration is one of the most important problems, which affects synthesis of the virtual and actual environments. Its main challenge lies in accurate alignment of the virtual objects in the real environment, with AR display properly displayed with respect to the real objects.

As described in Chapter 2, current registration methods in IGS can be divided into two types: sensor based and computer vision based registration. In this study, we especially focus on computer vision based method that can fully take advantage of the ProCam system to build an effective registration in a SAR environment. The proposed registration method uses a surface matching algorithm to find the optimally matched points between two data spaces, surgical model space and patient surface. This surface matching registration is used to align the surgical models (anatomical models and surgical planning models) with their corresponding spatial positions on the patient body. In order to properly deploy the projection of these surgical models on the patient (skin) surface, a point-based registration method is developed to align the model projection with its corresponding organs.

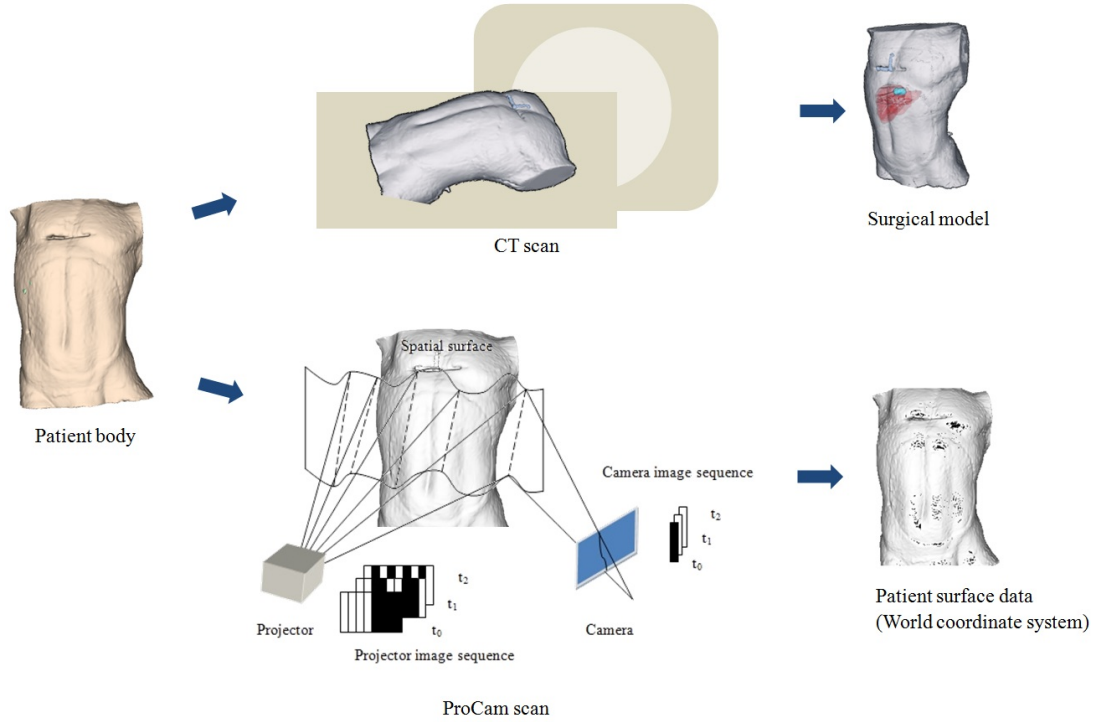


Figure 5-1. Data acquisition for registration.

5.1 Registration between Surgical Model and Patient Body

5.1.1 Data Acquisition and Preprocessing

In this SAR based surgical navigation system, the data sets used for registration between the virtual and real scene are acquired from the CT scan and ProCam scan respectively (Figure 5-1). The data spaces related to ProCam AR environment are listed in Table 5.1. The registration procedure involves spatial data transformation among the projector, camera, patient and the model spaces.

The data for virtual scene construction consists of patient-specific medical data including pathological and anatomical information, and surgical preplanned data. The patient-specific medical data can be acquired by CT scan as shown in Figure

5-1.

Table 5.1: Data spaces used for ProCam-baed AR construction

| Symbol | Name | Notation |
|----------------|------------------------|--|
| \mathbf{W} | World space | Based on the coordinate system constructed by the stereo vision device, originating at the principle point of the left camera. |
| \mathbf{P}_t | Patient space | Based on the coordinate system constructed by the fiducial markers attached to the patient. |
| \mathbf{M}_a | Anatomical model space | Based on the coordinate system of computer-generated models, which are constructed from the CT slices. |
| \mathbf{M}_b | Ablation model space | Based on the coordinate system of the ablation models. |
| \mathbf{C} | Camera image | Based on the coordinate system of the camera screen. |
| \mathbf{P} | Projector image | Based on the coordinate system of the projector screen. |
| \mathbf{R}_b | Robot system | Based on the coordinate system of fiducial markers attached to the surgical robot. |

The transformation from workspace X to Y is denoted as ${}^Y\mathbf{T}_X$.

A noninvasive marker frame is also scanned with the patient body during this procedure. The marker frame is used to construct a coordinate system to

register the patient model space with the world space. Patient's slight movement (e.g. free breathing) can also be tracked by observing the markers with the stereo vision system. With the CT-scan data, the critical anatomical areas in the CT image sequence, such as artery, hepatic vessels as well as the tumor areas, are then segmented by a hybrid manual and automatic method (Wang et al., 2010a). The anatomical models and tumor model can be constructed from the segmented profiles.

In order to establish a datum reference in the real environment for the virtual scene construction, the patient body is scanned by the ProCam system to acquire its surface data in the world coordinate system (Figure 5-1). In this case, the ProCam system can be considered as an integrated system for both AR construction and real-scene data acquisition. The geometry of retrieving the patient surface data with the ProCam system is shown in Figure 5-2.

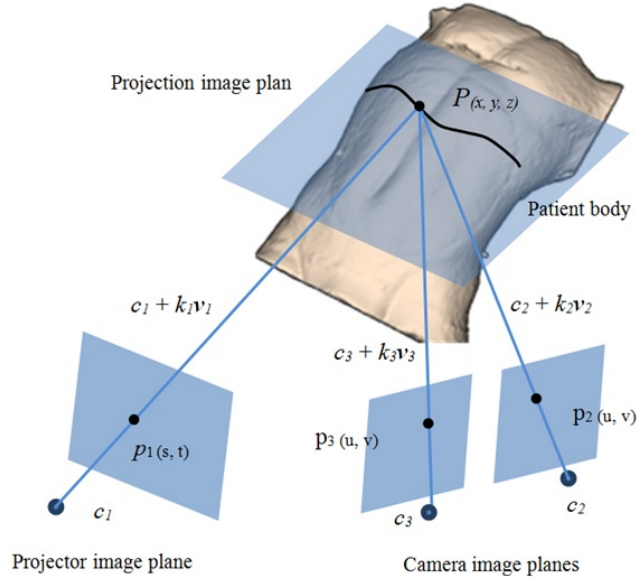


Figure 5-2. Geometry for retrieving the surface data with a projector-camera system.

In Figure 5-2, the image correspondence $p_{1(s,t)}$ and $p_{2(u,v)}$, $p_{1(s,t)}$ and $p_{3(u,v)}$ can be acquired by the method of structured-light illumination which is described in Chapter 3 (Wen et al., 2010b). Assume that the ProCam system is calibrated

and the coordinates of the projector and camera can be acquired. Under this assumption, the equations of projected planes and rays, as well as the equations of camera rays corresponding to illuminated points, are defined by parameters which can be measured (Lanman and Taubin, 2009). From these measurements, the location of illuminated points can be recovered by intersecting the planes or rays of light with the camera rays corresponding to the illuminated points. Through such procedures, the depth of a spatial point on the object (patient) surface can be recovered.

While the intersection of a ray of light with the object being scanned can be considered as a single illuminated point, the intersection of a plane of light with the object generally contains many illuminated curved segments (Figure 5-2). Each of these segments is composed of many illuminated points. A single illuminated point, visible to the camera, defines a camera ray. The vector \mathbf{v}_1 is the ray illuminating the surface point $P_{(x,y,z)}$ from the point $p_{1(s,t)}$ on the projector image, while \mathbf{v}_2 and \mathbf{v}_3 are the ray vectors reflected from $P_{(x,y,z)}$ to the points on the two image planes of the stereo vision device respectively. These two pairs of vectors intersect at $P_{(x,y,z)}$ in the world coordinate system. The point $P_{(x,y,z)}$ on the patient surface is then derived by searching the minimum distance between the points lying on the two ray vectors \mathbf{v}_1 and \mathbf{v}_2 , and \mathbf{v}_1 and \mathbf{v}_3 ,

$$\phi(k_1, k_2) = \|(c_1 + k_1\mathbf{v}_1) - (c_{2,3} + k_{2,3}\mathbf{v}_{2,3})\|^2, \quad (5.1)$$

where c_1, c_2, c_3 are the principle points of the projector and cameras respectively. k_1 is scale factor for \mathbf{v}_1 , and k_2 and k_3 for $P_{(x,y,z)}$ to reach $p_{2(u,v)}$ and $p_{3(u,v)}$. Based on this line-line intersection, the coordinate of the surface point $P_{(x,y,z)}$ can then be retrieved from the vectors $\mathbf{v}_2, \mathbf{v}_3$.

Once the data space of actual patient surface \mathbf{P}_t and the surgical model \mathbf{M}_a

are constructed, their registration can be achieved by surface matching algorithm which minimizes their l_2 -norm distance for point clouds optimal matching.

5.1.2 Surface Matching for Optimal Data Alignment

Assume p_i^m , $\{p_i^m = (x_i^m, y_i^m, z_i^m)^T | p_i^m \in \mathbf{M}_a\}$, and p_j^s , $\{p_j^s = (x_j^s, y_j^s, z_j^s)^T | p_j^s \in \mathbf{P}_t\}$, where $i = 1, 2, \dots, M$, $j = 1, 2, \dots, N$, are the coordinates of the patient surface points that are derived from CT and ProCam scanning respectively. To find the correspondence between these two data sets, normalization of each data with their respective marker frame as a reference is initially performed. Since the CT scanning usually generates more dense data, $M > N$, the correspondence of point p_j^s in \mathbf{M}_a can be estimated by filtering the points with distance $\sqrt{(p_i^m - p_j^s)^2} \geq D$, and then finding the nearest point around p_j^s .

The relationship between the two data spaces can be expressed by the following spatial transformation in terms of the rotation matrix \mathbf{R} and translation matrix \mathbf{T} ,

$$p_i^m = \mathbf{R}p_j^s + \mathbf{T} + \zeta_i, \quad (5.2)$$

where ζ_i is the noise data (x_i^n, y_i^n, z_i^n) between the two data spaces. This may arise from different accuracy levels of data acquisition with the ProCam system and CT scanner. Uncertain factors such as lightings variation in the working environment may also contribute to ζ_i .

Suppose another data set in space \mathbf{M}_a is p_i^e , $\{p_i^e = (x_i^e, y_i^e, z_i^e)^T | p_i^e \in \mathbf{P}_t\}$, and $p_i^e = \hat{\mathbf{R}}p_m^i$. It has been proven that if $\hat{\mathbf{R}}$ is the least-square solution to Equation 5.2, p_i^m and p_i^e have the same centroid (Arun et al., 1987), $p_o^e = p_o^i$. p_o^e and p_o^i are the centroid of their data sets respectively.

In order to find the estimated the rotation transformation $\hat{\mathbf{R}}$ based on the same

centroid, the two data sets p_i^m and p_i^e are normalized firstly by Equation 5.3 and 5.4,

$$\Delta p^m = p_i^m - p_o^m, \quad (5.3)$$

$$\Delta p^e = p_i^e - p_o^e. \quad (5.4)$$

$\hat{\mathbf{R}}$ can be estimated by minimizing φ ,

$$\varphi = \sum_{i=1}^N \|\Delta p^m - \mathbf{R} \Delta p^e\|^2. \quad (5.5)$$

To solve \mathbf{R} in Equation 5.5, singular value decomposition (Hartley and Zisserman (2003)) can be used. By decomposing \mathbf{H} ,

$$\mathbf{H} = \sum_{i=1}^N \Delta p^m \Delta p^{eT}, \quad (5.6)$$

$$\mathbf{H} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}. \quad (5.7)$$

In this case,

$$\hat{\mathbf{R}} = \mathbf{V} \mathbf{U}^T. \quad (5.8)$$

With the estimated rotation matrix $\hat{\mathbf{R}}$, the estimated translation matrix $\hat{\mathbf{T}}$ is then acquired by

$$\hat{\mathbf{T}} = p_i^m - \hat{\mathbf{R}} p_j^s. \quad (5.9)$$

Figure 5-3 shows the registration with surface matching algorithm on a mannequin body. The blue regions are the registered point cloud of the data sets acquired by CT and ProCam scanning respectively.

In order to accelerate registration between the surgical model and patient body, fiducial markers are used to dynamically register the anatomical model space with the world space. It enables the virtual camera in the model space to register

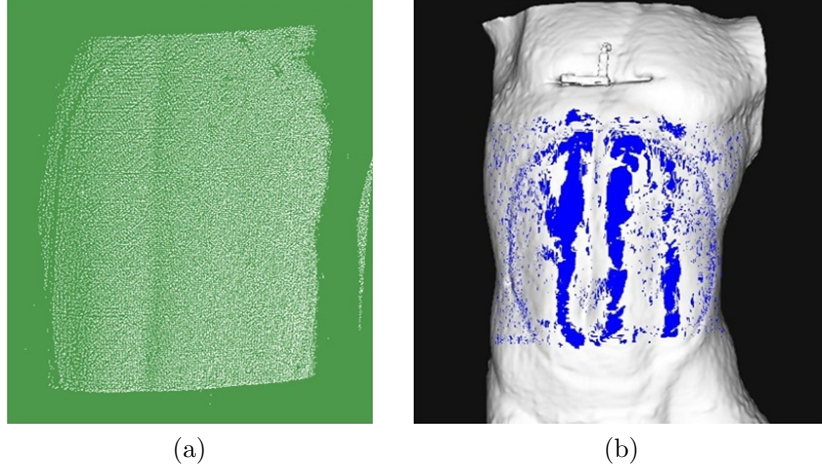


Figure 5-3. Surface matching based registration between patient's surface point cloud (a) and surface model (b). The blue regions represent the matched surface data.

with the corresponding actual camera of the stereo vision system in the world coordinate system. Additionally, this point-based registration simplifies update of the needle position in the model space during surgery. In the anatomical model space, the virtual camera's position with respect to the fiducial marker model is registered by

$$\mathbf{T}_M^{m-c} = s\mathbf{T}_a\mathbf{T}_{reg}^{\mathbf{P}^t}\mathbf{T}_c, \quad (5.10)$$

where $\mathbf{P}^t\mathbf{T}_C$ is the transformation from the stereo vision device to the fiducial marker frame in the world space (Table 5.1), \mathbf{T}_{reg} is the model-patient registration matrix, $\mathbf{T}_{reg} = \hat{\mathbf{R}}\hat{\mathbf{T}}$. s is a derived scale factor, and \mathbf{T}_a is an affine transformation matrix that registers the viewing angle in the model space with the principle axis of the stereo vision system in the world coordinate system.

Finally, with the surgical model-patient registration matrix \mathbf{T}_{reg} , the anatomical and ablation models constructed in the preoperative planning procedure are registered with its corresponding spatial positions in the patient body. The pre-planned insertion points and trajectories are transferred into the current patient space following the model-patient registration. The point-based registration with

the fiducial markers enables registration of the position of the actual camera with the virtual camera in the anatomic model space \mathbf{M}_a .

5.2 Registration between Model-Projection Image and Patient Body

Registration between projection model and patient body aims to align the model projection with its corresponding actual objects in the world coordinate system. As described in Section 5.1, the marker frame provides a spatial relationship between the anatomical model space and the patient space. Point-based registration method with light-emitting diode (LED) marker frame has been investigated in an robot-assisted AR surgery (Wen et al., 2010a). With the markers attached onto the patient's skin, spatial relationship between the markers and target organ can be derived from CT image of the patient.

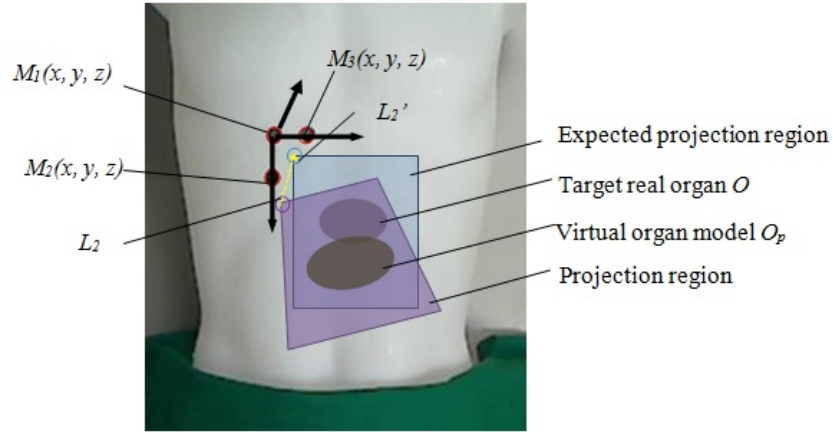


Figure 5-4. Marker-based registration for SAR (M_1, M_2, M_3 are three markers attached onto the mannequin body.)

As shown in Figure 5-4, three markers M_1, M_2, M_3 are attached onto the patient body to construct a space reference frame T . This marker frame is used to construct the spatial relationship that connects the stereo vision frame, world coordinate and CT image frame. The spatial relationship enables the virtual model displayed by

the ProCam system to register with the target organ within the patient body. The two steps are described as follow to overlay a virtual model of the target organ properly on the patient with correct position, scale and angle.

The first step involves establishing the spatial relationship between the markers M_i ($i = 1, 2, 3$) and the target organ O . CT imaging of the patient with three markers attached is performed prior to the surgery. Appropriate markers deployment is important. The markers should facilitate the registration and avoid inconveniencing the surgery. Their deployment is dependent on the following considerations. Firstly, the surgical tool's motion should not be obstructed by the markers. The surgical path or motion is given a higher priority in preoperative planning compared to that of marker's deployment. Since the markers are used for constructing the space reference frame T , this can be achieved as long as the stereo vision system can detect the markers during the registration. Preoperative planning including obstruction prevention will be discussed in Chapter 6. Secondly, the markers should be placed beyond the expected projection areas to avoid violating the projected virtual models.

The initial ProCam system calibration described in Chapter 3 and 4 has already established pixel correspondence between the projector and camera. The correspondence works as a lookup table which can define any corresponding regions between the projector and camera images. With the stereo vision system, the reference frame T constructed by three markers is achieved from the estimation of space coordinate of the three markers. As described above, the spatial relationship between the markers and the target organ can be derived from the CT images. In this case, the spatial relationship between target organ and the projected virtual model is established by

$$R_O^T = R_{m_i}^T \times R_o^{m_i}, \quad (5.11)$$

where R_i^j represents the spatial transformation from point i to j . Since the coordinate of the top-left corner of the projection image L_2 is easily detected, the position of L_2 can be adjusted to its expected coordinate L'_2 . Assume that the L_1 and L_3 are the top-left corner of the source image and captured image respectively, and L'_1 and L'_3 are the expected position of the top-left corner of the source image and captured image respectively. The expected coordinates of the projected image can be acquired by calculating the spatial relationship with the homography between the camera and world coordinate. The relationship between $L'_1(x, y)$ and $L'_3(x, y)$ is described by

$$L_3 = \mathbf{M}_A L'_3, \quad (5.12)$$

$$L'_1 = \mathbf{H}_p \mathbf{H}_c^{-1} L'_3, \quad (5.13)$$

where \mathbf{M}_A is a affine transformation. A testing projection is conducted to detect the distance between L_3 and L'_3 . With the piecewise linear method described in section 4.1, \mathbf{M}_A and $\mathbf{H}_p \mathbf{H}_c^{-1}$ can be replaced by the $\mathbf{H}_{p_i} \mathbf{H}_{c_i}^{-1}$ and \mathbf{M}_{A_i} for corresponding piecewise region.

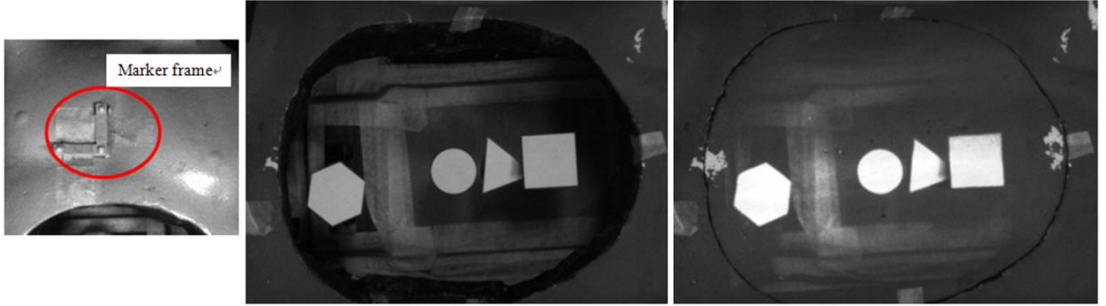


Figure 5-5. Geometric correction and registration of the model-projection on an irregular surface with its corresponding internal object.

A marker-frame guided registration that aligns model-projection of inner objects with their corresponding position on the surface is shown in Figure 5-5. It demonstrates that this point-based registration method can register the projection on an irregular surface without sacrificing geometry accuracy. Assume that the

patient lies on a surgical bed and the stereo camera is placed vertically above the target organ. A virtual organ model with the same size and angle located as the real one is then expected to be projected on the patient body. This bird view of the virtual model display can be derived from the CT images. Finally, geometric correction, surface matching algorithm and point-based registration method are integrated to register the virtual models with their corresponding real target organs.

5.3 Summary

Chapter 5 presents our achievements on registration methods for construction of the spatial AR in surgical environment. It solves registration problems between the preplanned surgical models and patient, and registration between surgical model projection image and the patient. The data spaces related to this projector-based AR guidance system are listed in Table 5.1.

With the ProCam system, we examine line-line model for retrieving spatial points from an object patient surface. The phase-shift gray-coded patterns can be used to acquire the ProCam pixel correspondence. In particular, we propose a surface matching based algorithm to register preplanned surgical models including patient anatomy models with their corresponding positions on the patient surface. The combination of marker-based registration with the surface matching algorithm increases accuracy and registration efficiency. The effectiveness is validated on an *in vivo* and *ex vivo* experiment (Chapter 7). One advantage of this method is simplicity: only a marker frame with three points needs to be specified at the pre-diagnostic imaging stage. Another advantage of the proposed method is that it takes full advantage of the ProCam system as an integrated system which can be used for acquisition of surface data from patient body and construction of a spatial

AR environment simultaneously. In this case, the patient (skin) surface data can be acquired during the surgery and the virtual objects can then be registered on the patient body.

As to registration between the model-projection image and patient body, a point-based registration method is proposed. LED markers are used to construct the spatial relationship among the CT images, real organs, ProCam system and the world coordinate system. With the spatial relationship, the initial model projection can be adjusted to its corresponding organ areas.

Chapter 6

AUGMENTED INTERACTION

In the previous chapters, we discussed the construction of ProCam-based SAR environment that enables computer-generated graphical models to properly overlay the real world without geometric and radiometric distortion. All those parts could be considered as output technologies of AR. In this chapter, the relationship between the user input and its corresponding output is presented, and the methods of augmented interaction for surgical environment are studied.

While AR solutions have mostly focused on output technologies such as HMD and hand-held displays, most of them suffer from inconsistent viewing focus, limited field of view or encumbered ergonomics. For example, in order to interact with virtual content, users have to wear sensors and trackers, HMD displays, or required to set up a semi-transparent mirror to guide the interaction (Benko et al., 2012). The current interaction modes may also be limited in certain applications. In a surgical environment, there are high requirements on interaction manner (direct and intuitive interaction), sterilization (non-contact operation) and ergonomics (few cables attached) which are described in Chapter 2.

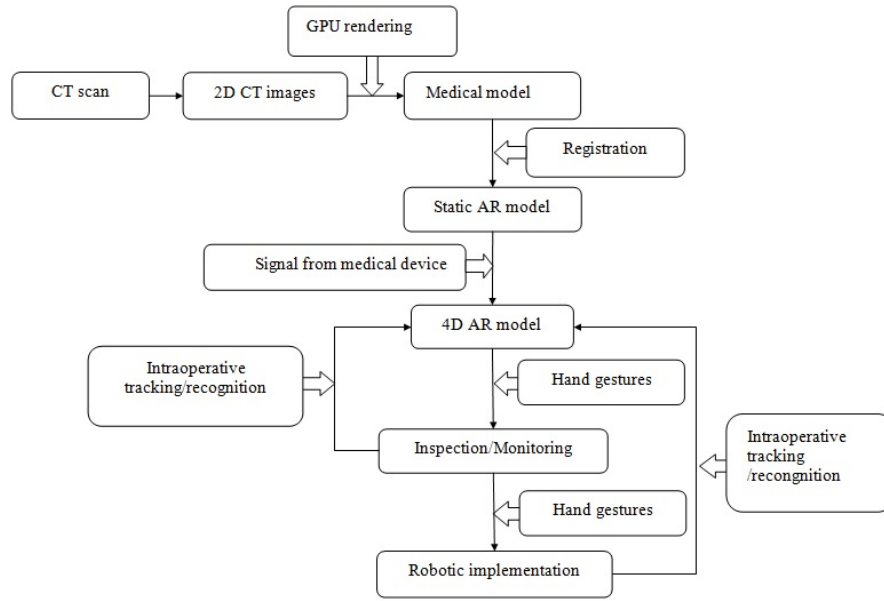


Figure 6-1. Work flow of the proposed interface for an AR-guided surgery.

The general work flow using the ProCam system for HCI in a surgical environment is illustrated in Figure 6-1 (Wen et al., 2010a). It firstly involves the handling of 3D medical models that are reconstructed from CT images and preoperative surgical planning. Then there is an iterative process with real-time AR model rendering and update of the intraoperative information from the surgical field, which includes hand gesture interaction and surgical robot implementation. When AR projection field receives "stimulus" from the real environment, the tracking and recognition modules will respond to the stimulus information to update the states of AR models, which provides necessary visual cues to guide surgeons' decision and operation.

6.1 Preoperative Planning

Radiofrequency ablation (RFA) is used for treatment of the unresectable liver malignancies with advantages of safety, ease of application and thermal ablative consistency (Garrean et al., 2008). Radiofrequency generator produces high-frequency

(200-1200 Hz) heat on the tips of the needle electrodes to create a region of necrosis of local cellular death. Although RFA is an effective approach for treatment of the liver metastatic, it faces challenges of inadequate visual information in RF needle navigation and handicaps in operation dexterity. In recent years, development of the medical imaging techniques assists surgeons to identify the preoperative surgical models and intraoperative guidance information in surgery.

In the construction of an RFA surgical model, users are allowed to design their own models in which they can define different trackers to intraoperatively guide the projection display. These trackers are with large range of parameters for graphic rendering to adapt to user's input. In the RFA surgery, interactive planning involves ablation model planning and surgeon's supervisory feedback via ProCam guidance. An optimally preplanned ablation model is needed to ensure the effective ablation of the whole tumor with the smallest number of insertions. The preoperatively defined surgical planning data is visualized directly on the patient body through the AR display supplying information of insertion points, ablation points and preplanned trajectories.

A systematic surgical planning is a necessary preparation with the patient-specific medical data to generate the optimal ablation points, insertion points and insertion trajectories for the subsequent robot-assisted RF treatment. These surgical models and data are integrated into surgical AR environment and used to guide surgeons' intraoperative supervision. Surgical treatment of liver tumor typically starts with the acquisition of relevant pathological information and patient specific anatomical knowledge through the diagnostic imaging modalities (CT, MRI, Ultrasound). In this study, the medical data including pathological and anatomical information is acquired through patient CT scanning. The critical anatomical areas on the acquired CT images such as artery, hepatic vessels as well

as the tumor areas are segmented by a hybrid manual and automatic method. In this case, the anatomical models and tumor model can be constructed from the segmented profiles. A noninvasive marker frame attached on the patient's skin is also scanned for this preplanning.

In order to efficiently process tumors' 3D profiles and irregular shapes, our group has proposed a voxel-growing algorithm (Yang et al. 2010) to automatically produce the ablation points which are defined as the trackers. "Growing regions" are optimally aligned to best fit geometry of the target tumor areas, which reduces the number of insertion times and minimizes the burning areas of healthy tissue. The input for ablation model is the tumor's surface data, which is acquired from CT scan in advance. Our method involves the construction of spherical elements defined by its inscribed cubes. Each sphere represents the ablation area, and the inscribed cube represents a voxel element. Voxel elements propagate layer by layer along x, y and z axis to cover the entire tumor area. With voxel elements propagating, the coordinates of the voxel elements' centers are produced, which generates ablation coordinates (Figure 6-2). Details of the algorithm are presented as follows.

Step 1: Determine the size of a voxel element. The size of a voxel element can be defined by specifying the dimension of a single ablation. The maximum ablation size is used to minimize the required number of ablations. For most RF device, this value is 5 cm (Chen et al., 2004).

Step 2: Compute required layers along z axis. Assume the z coordinates of the surface data which are obtained from CT scan are $s[1][z], \dots, s[i][z], \dots, s[n][z]$, where i is the i^{th} coordinate of the surface data which expressed in 3D coordinate, and z is the z coordinate. In this way, we can compute the maximum and minimum z among the surface data.

Step 3: The voxel element propagates along x and y axis in each layer which is constructed along z axis. When voxel element propagates to the margin of the tumor area, boundary condition is examined to make sure the ablation volume covers the whole tumor area with minimum burning of peripheral healthy tissues.

This planning algorithm is based on clinical optimizing conditions and practices. We adapt our planning algorithm based on the cylindrical model discussed in (Dodd et al., 2001). As such, this model adopts uniform spherical ablation zone for linear propagation. However, the configurations of these linear rows are "packed" in an optimal fashion to compromise the irregularity of the tumor geometry. The utmost priority of this treatment plan is to ensure complete ablation of the entire tumor because leaving any tumor cell untreated can lead to recurrence and further spreading of the tumor. Thus, our planning strategy skewed conservatively towards complete ablation coverage of target rather than attempting optimal control of ablation geometry.

Despite being in agreement with many criterias defined in the existing literature (Baegert et al., 2007)(Seitel et al., 2011), we recognize that there are other clinical considerations and case-specific prioritization of criteria. For instance, the surgeons might consider prioritizing conservation of hepatic reserve over operation duration and minimizing needle punctures. In such situation, the outer most voxel elements exposed to healthy tissue can be further subdivided to refine the contour conformity of the model ablation region. There is generally no one-size-fits-all solution for the best ablation model. The voxel-growing based method is executable with simple manipulator trajectories yet highly adaptive to various clinical cases. The disadvantage of this method might be the expense of more needle punctures and longer operation duration.

With the above voxel-growing algorithm, a preplanned ablation model can be

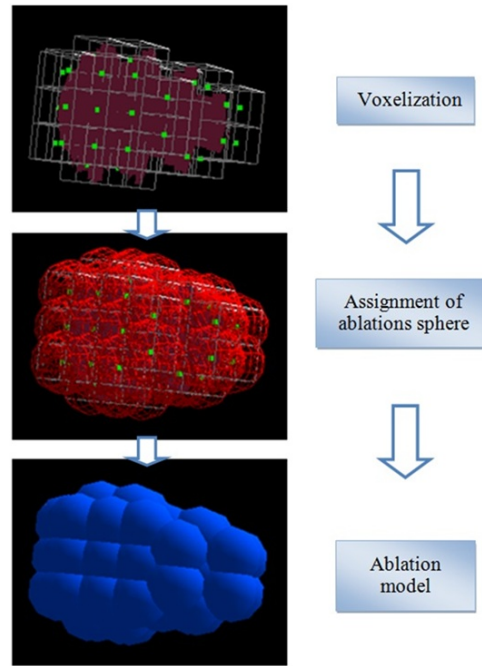


Figure 6-2. Construction of the optimal ablation model (brown figure: virtual construction of tumor; green dots: designated location of needle tips; red wireframe: predicted ablation region; blue: resultant necrosis region).

derived from tumor's 3D profile. The growing elements propagate along the x, y and z axis to automatically generate ablation points within the tumor model's coordinate system. A safety margin of 10 mm is considered into the ablation model to comply with the standard clinical practice (Wen et al., 2010b). The margin can be adjusted by selecting different growing elements, which results in different accuracy levels. With the generated ablation points, the insertion trajectories can be generated from the insertion points located on the patient skin to the ablation points with respect to tumor's peripheral anatomical structures (Figure 6-3a).

To ensure the anatomic structures are not harmed and the bones are avoided by the RF needle, the needle's position and orientation could be optimized by a gradient-based method (Du and Zhang, 2008). Based on the above mathematic ablation model and insertion trajectories, surgeon's professional expertise and surgical experience are expected to examine the planning feasibility and its potential

risks. In addition, the workspace of the surgical robot with respect to the pre-planned needle placements is taken into consideration. It covers the entire tumor region and all the insertion trajectories avoiding the fiducial markers attached on the patient (Figure 6-3b).

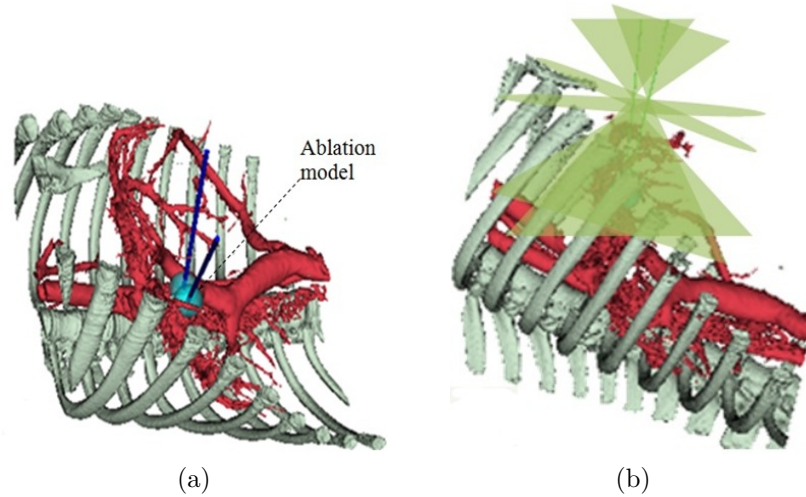


Figure 6-3. Surgical model (ablation model) based surgical planning: (a) ablation model planning is based on anatomic models; (b) path planning is based on available workspace of the surgical robot.

6.2 Interactive Supervisory Guidance

The objective of augmented interaction design is to develop an effective mechanism that integrates preoperative surgical planning, surgeon's supervisory guidance and surgical robot-assist implementation for MIS. Contrary to the concept of a fully-automatic surgical robot system, we consider the surgeon's professional expertise and experience as too essential to be excluded to guarantee a safe and efficient surgery. In this case, the proposed mechanism encompasses three modules: a projector-based AR display for visualization of the planning data directly on the patient body, surgeons' interactive supervision, and a robot-assist execution following the pre-surgical plan and surgeon's guidance.

In the intraoperative interaction, hand gesture is adopted to control display of the AR model and augmented needle insertion. Surgeons' hand gestures are detected by a stereo vision system in the AR environment. This interface has the advantage in sterilization which is an important requirement during operating. The vision based hand gesture can be directly integrated into the AR system using a ProCam system or conventional vision based tracking device. The AR model projected on the patient surface can be scaled and rotated with hand gesture guidance, which facilitates intraoperative insertion check. Additionally, hand gestures can also be used to guide robotic execution.

In order to enhance the speed and accuracy of hand gesture recognition, a 3D model based hand gesture recognition method is proposed. This method is based on the analysis of model features produced by a space palm plan as well as 3D convex hull points. Users are allowed to pose the defined hand gestures without attaching markers or data glove. Due to human hand's flexibility, stereo vision system is used to resolve ambiguity on recognition of the hand features. As shown in Figure 6-4, the 3D model is constructed including the space palm plan which is relatively stable, and a brace model of five fingers with the key joints and finger tips. Before starting the hand gestures recognition, the 3D hand model is initialized by characterizing the model with the specific hand parameters such as palm size, fingers layout of different users. This initialization can be achieved by directly measuring the key parameters of the user's hand under the vision system. These key parameters include hand length and width, the ratio between the palm and the fingers, and the palm's location relative to the stereo vision system. Palm plan can be acquired by regression of the hand palm coordinates as shown in Figure 6-4(b). There are several ways of tracking the user's hands (Wen et al., 2010a) (Wen et al., 2010b). In order to enhance robustness of the introspective recognition, the background of hand operation in the surgical AR

environment is covered by the surgical cloth which is usually a green background. CamShift method (Wen et al., 2010a) used for medical instrument tracking can also be integrated with hand tracking. When the hand tracking is implemented, the palm plan is constructed in real time to update its pose in the world coordinate system. Following with the hand tracking, the 3D hand model is attached to the user's hand wherever the hand moves. In this case, parameters of the 3D hand model are analyzed and the hand gesture recognition is achieved in real-time.

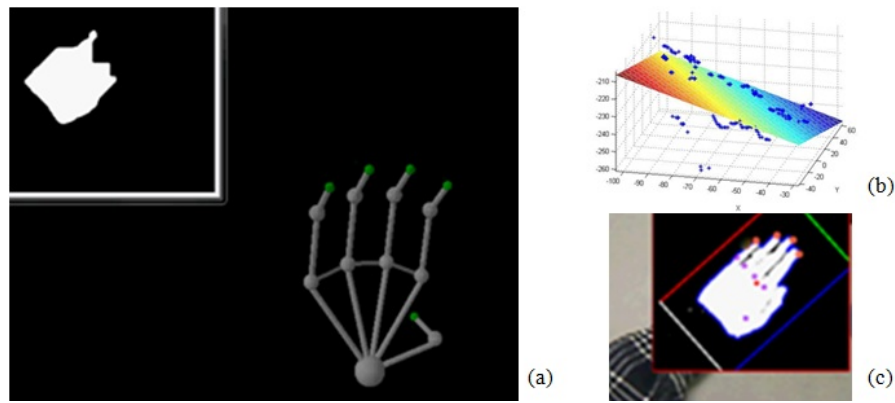


Figure 6-4. 3D model based hand gesture recognition. (a) 3D graphic model of hand gesture and its corresponding 2D processed image. (b) Hand plane derived from the spatial point cloud of a hand gesture. (c) Key geometric parameters of the hand gestures

The whole process of gesture recognition is divided into stages of image pre-processing, training and prediction. The input images are firstly pre-processed to extract the features and key parameters mentioned above. The Naive Bayes Classifier is then used to enhance computational efficiency because less training data is involved (Al-Aidaroos et al., 2010).

Image Pre-processing: The hand gesture is captured by a stereo camera system at the rate of 61 f/s with a resolution of 680 x 480. Each frame is analyzed as a single input image. In order to segregate the hand from the background, skin colour detection is used. The colour model of the image is converted from the default RGB model to YCbCr model to allow separation of the luma component from chroma component of the color space (Majumder and Brown, 2007). Skin

color detection of can then be done at varying levels of light intensity, which is difficult to achieve if the RGB colour model is used instead. The contour of the hand is then extracted using 3D convex hull. Using the contour points for the left and right bounds, the input image is partitioned into a fixed number of vertical segments and a frequency histogram for the hand pixels is constructed. This data is then used to train the classifier model. The capturing of the frames from the camera, the background segmentation and feature extraction are done with multi-threading to exploit parallelism and accelerate the process.

Training: A training data set of 1400 sample images for all the gestures is used to train the 3D gesture models. A training size of less than 100 images per gesture does not yield any reasonable results, thus at least 300 images are used to train each gesture.

Prediction: The prediction by the classifier is only registered as a positive result if the output class is consistent for at least 10 consecutive frames to reduce the number of false positives during transitional motions.

Evaluation: Preliminary tests showed that the system can recognize the gesture to a reasonably high degree of accuracy, but only with appropriate environmental conditions such as adequate lighting and discernible background color. The contour points are initially used solely as the feature for recognition, but this does not yield any reasonable results. Thus a decision was made to use pixel histogram instead, which provided better accuracy and invariance of hand size, distance from the camera and horizontal position of the hand. This makes the system more robust to different users and environmental conditions. The technique assumes the hand is the dominant skin colored object in the frame. It is also dependent on the orientation or angle of the gesture with respect to the frame as the pixel histogram will vary at different orientations.

6.3 Augmented Needle Insertion

One of the outstanding advantages of ProCam-based AR is it has large field of view and can directly incorporate actual tools interaction with virtual objects. This section presents augmented interaction with robot-assisted needle insertion for ProCam system based surgical guidance (Figure 6-5a).

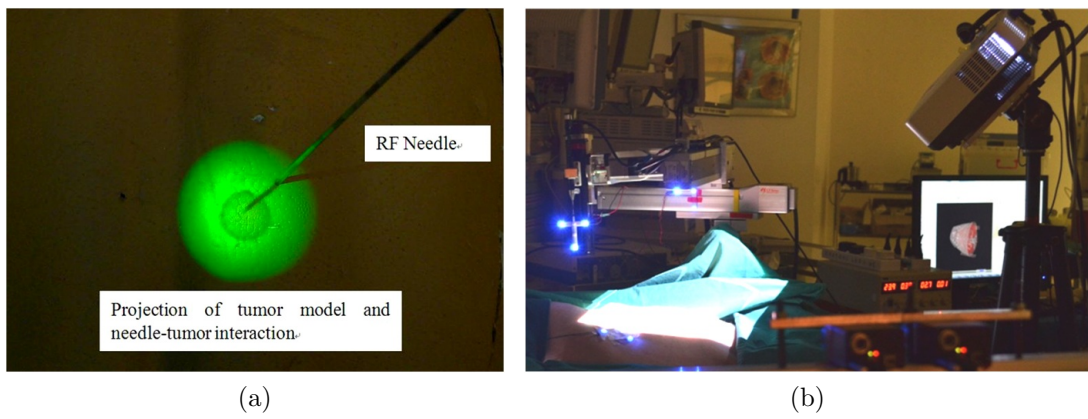


Figure 6-5. (a) ProCam-based augmented needle insertion (b) Initialization of the surgical robotic system in an operating room.

In the initialization of the robotic system, the RF needle's position is calibrated via the surgical robot system calibration module (Yang et al., 2010) (Figure 6-5b). LED marker frames are attached on the needle's shaft, robot arms and base to facilitate the calibration. With vision-based detection of the LED markers by the stereo vision system, the parameters of the surgical robot are measured and the coordinate system of the robotic system is established. Based on the needle length and the spatial data of the LED marker frame that is attached on needle shaft, the needle's position including coordinates of the needle tip within the robot coordinate system can then be calculated. The needle's spatial insertion (actual insertion trajectory) is intraoperatively navigated by stereoscopic tracking of the LED markers, and needle-tissue simulation between the anatomic and RF needle models (Figure 6-5a). The robotic execution benefits from improved accuracy on

needle's tip tracking and model-based simulation.

The previous model-patient registration in Chapter 5 enables preoperative planning data to be mapped into patient space and world space. Visualization of augmented needle insertion can assist surgeons to intraoperatively detect whether the needle trajectory is following the preplanned one within the patient body. In addition, the tumor model and ablation points can be directly checked (e.g. rotate, zoom in and zoom out) by hand gesture operation. The display is achieved by model simulation and projection which are updated by the ProCam system in real time.

Transformation between the different work spaces (Table 5.1) for intraoperative visual guidance is illustrated in Figure 6-6. The surgeon's viewing position with respect the stereo vision device can be tracked in real time (Garstka and Gabriele, 2011).

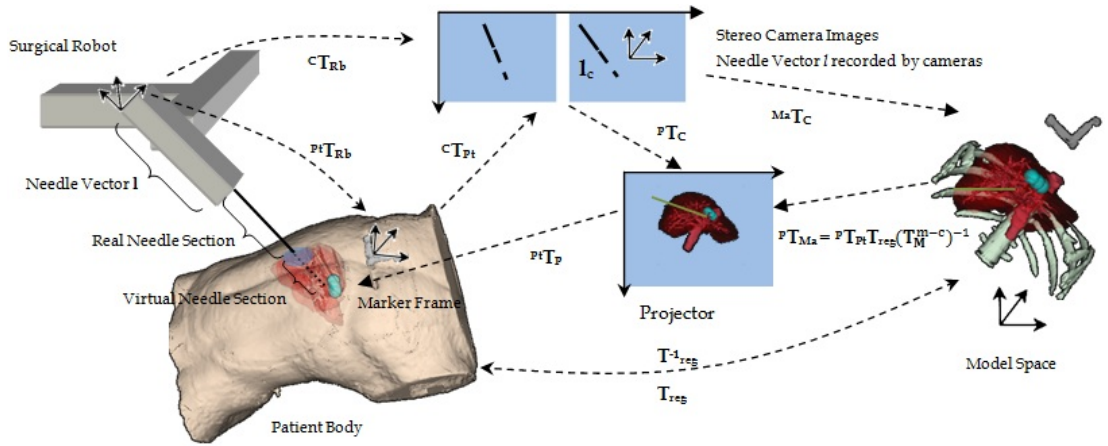


Figure 6-6. Transformation between the different workspaces for intraoperative augmented needle insertion.

Each preplanned insertion point p_i^{in} is visually labeled by projecting a crossing mark on the patient surface. The transformation from p_i^{in} to its corresponding projection point p_i^{ins} on the projector image is described by

$$\mathbf{p}_i^{ins} = {}^{\mathbf{P}}\mathbf{T}_{\mathbf{P}_t}\mathbf{T}_{reg}p_i^{in}, \quad (6.1)$$

Similarly, the transformation of the needle placements along the preplanned trajectories to their corresponding points displayed on the patient surface also follows the transformation matrices in Equation 6.1. The transformation from the RF needle's current position p_i^{rb} to the preplanned insertion points p_i^{in} is given by

$$p_i^{in} = \mathbf{T}_{reg}^{-1}{}^{\mathbf{P}_t}\mathbf{T}_{\mathbf{R}_b}p_i^{rb}, \quad (6.2)$$

When the RF needle approaches the preplanned insertion points, the parameters of needle's motion and pose (inserting angles and position) are adjusted under the guidance of stereo vision device. The needle vector is monitored to ensure that the vector angles with respect to the fiducial marker frames on the patient body, α, β, γ , are within the satisfactory range ($\theta_1 \leq \alpha, \beta, \gamma \leq \theta_2$).

In order to display the current needle's position inside the body during insertion, a projection of the partial needle model that is consistent with the real needle section outside the body is obtained (Figure 6-6). The preplanned trajectory is incorporated into algorithm EKF as a reference to search for the needle tip in patient space. We assume that the needle is a rigid body and does not bend during the insertion procedure. The needle's vector in the world space is presented as a spatial line vector \mathbf{l} (Figure 6-6). It has been proved that straightness is preserved in the geometry of projective transformation (Hartley and Zisserman, 2003). Thus, the perspective transformation of the needle vector among the working spaces is a line vector. This means the projection of RF needle's insertion part is consistent with the needle's line vector from camera's view. The transformation ${}^C\mathbf{T}_{\mathbf{P}_t}$ (Figure 6-6) assists construction of the line vector \mathbf{l}_c , $\mathbf{l}_c = (l_x, l_y, l_z)^T$, in the

camera image. Through the transformation described by

$$\mathbf{l}_p = {}^P\mathbf{T}_C {}^C\mathbf{T}_{P_t} \mathbf{T}_{reg} (\mathbf{T}_M^{m-c})^{-1} \mathbf{T}_{reg}^{-1} \mathbf{T}_{R_b} \mathbf{l}, \quad (6.3)$$

the RF needle's insertion part can be visualized by projecting image of the line vector \mathbf{l}_p . In this case, the line vector \mathbf{l}_c and \mathbf{l}_p are used as constraints to calculate the needle tip's position in the different workspaces. To dynamically predict the needle's position and pose (angles) in world space, the fiducial marker frame attached to the robot arm is tracked by the stereo vision device. The markers' coordinates and the z vector of the marker frame are estimated using the EKF:

$$\begin{cases} \chi_k = \chi_{k-1} + \xi_{k-1} \\ p_k = \mathbf{F}(\chi_k) + \mu_k \end{cases}, \quad (6.4)$$

where

$$F(\chi_k) = \begin{bmatrix} 1 & 0 & 0 & \frac{l_x - \cos \alpha_k}{\alpha_k} & 0 & 0 \\ 0 & 1 & 0 & 0 & \frac{l_y - \cos \beta_k}{\beta_k} & 0 \\ 0 & 0 & 1 & 0 & 0 & \frac{l_z - \cos \gamma_k}{\gamma_k} \end{bmatrix}, \quad (6.5)$$

the state $\chi_k = [x_k, y_k, z_k, \alpha, \beta, \gamma]^T$, (x_k, y_k, z_k) is the coordinate of the marker that lies at the crossing point of the marker frame, and k is the tracking time from the first needle insertion to the last one. ξ_k and μ_k are assumed to be sequences of white, zero-mean with constant covariance \mathbf{Q} and \mathbf{R} , respectively. P_k is the needle tip's spatial coordinate within the world coordinate system. l_x, l_y, l_z in $\mathbf{F}(\chi_k)$ are the x-, y- and z-component of the needle's line vector along the axes of the marker frame. The estimate $\hat{\chi}_{k|k}$ of the state χ_k can be obtained by

$$\hat{\chi}_{k|k} = \hat{\chi}_{k|k-1} + \mathbf{G}_k (p_k - \mathbf{F}(\chi_{k|k-1})), \quad (6.6)$$

where \mathbf{G}_k is the Kalman gain comprising the state prediction covariance and

Jacobian matrix of $\mathbf{F}(\chi_{k|k-1})$ which is used for nonlinear state estimation. Finally, the estimated needle tip position is generated by evaluating p_k .

In this study, we assumed that the RF needle is a rigid body and it keeps straight stiffness without any deflection during the insertion procedure. With this assumption, the actual depth of needle insertion is controlled by tracking the needle' position (position of the attached LED marker frame) to follow the preoperative surgical plan. The needle's depth and insertion angles are monitored and adjusted in real time to adapt to its preplanned insertion paths. The depth of needle insertion is pre-calculated for needle insertion during preoperative surgery planning and simulation. The projection-based AR guidance provides planned insertion points and virtual insertion paths on the patient body for needle placement guidance and surgeon's visual reference respectively.

However, in practice, the RF needle may be bending or deflected when it is interacting with soft inhomogeneous tissue within an organ. Recent studies (Tungjitkusolmun et al., 2002) on Finite Element (FE) model based needle model suggested a method to construct the bended needle model and deformed tissue model in real time during the insertion procedure. Other studies (Chui et al., 2004) (Chui et al., 2007) have discussed problems of needle deformation and soft needle positioning during the insertion procedure. However, these studies are beyond the scope of this study.

6.4 Summary

Chapter 6 is devoted to establishing interaction mechanism of the ProCam-based surgical AR guidance system. It includes hand gesture based direct human-AR interaction and augmented needle insertion between the surgical tool and AR

display. Preoperative planning is a necessary procedure to define AR display functions, rendering properties and interactive response. A method with 3D hand model is used to incorporate both hand features and spatial information to identify the user's input. The states of computer-generated models are then updated to adjust their AR display. This new method can realize an effective HCI mechanism that integrates the preoperative surgical planning and surgeon's supervisory guidance on surgical robot. With advantages of large field of view and direct user interface in ProCam AR display, a surgical robot assisted augmented needle insertion is explored. This may enable surgeons to overcome visual and operational limitations in a minimally invasive procedure.

Chapter 7

EXPERIMENTS AND DISCUSSION

In this chapter, we examine the previous methods and algorithms on *ex vivo* and *in vivo* porcine models with interactive AR display and guidance of RF needle insertion to validate the performance of ProCam AR system (Figure 7-1).

The ProCam AR guidance system, as shown in Figure 7-1, consists of an LCD projector, a stereo vision device comprising two high quality cameras, a surgical robot with 8-DOF, and a computer workstation equipped with quad-core processors, 12 GB memory and a graphics acceleration board supporting OpenGL 3D. The projector has a resolution of 1024×768 (XGA) and throw distance of 1.4 m 13.2 m. Resolution and frame rate of the cameras for the stereo vision device are 780×580 and 61 f/s respectively.

We experimented SAR display and augmented interaction on a mannequin ($85 \text{ cm} \times 55 \text{ cm} \times 23 \text{ cm}$) (Figure 7-2). Four plasticine models are implanted inside the mannequin body for accuracy evaluation of the projective AR construction,

model-patient registration and RF needle placement. The feasibility of RF needle insertion under this visual guidance was tested, and the needle placement along the actual trajectories was compared with the preplanned ones.

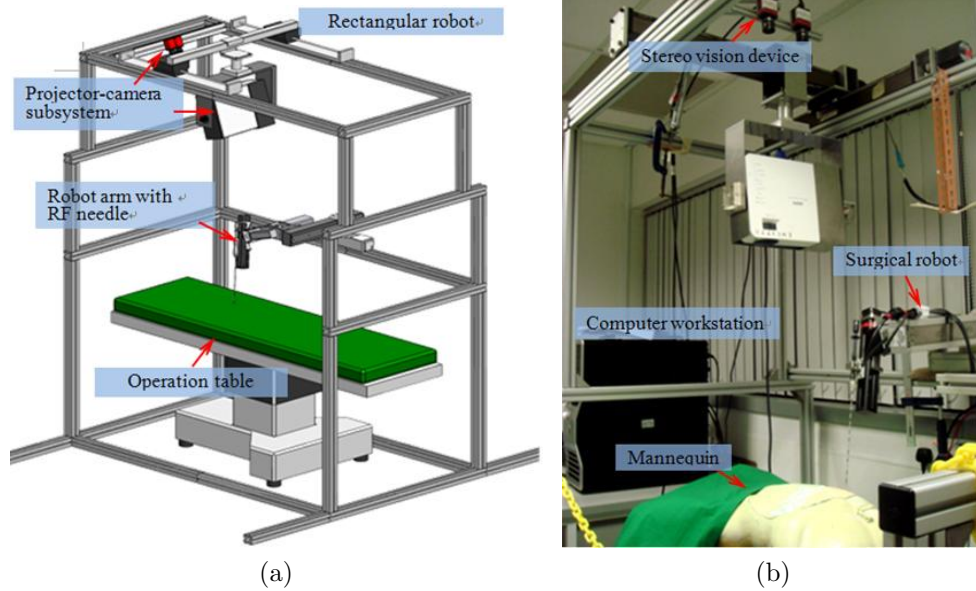


Figure 7-1. ProCam AR guidance system: (a) system model overview; (b) snapshot of the setup in the laboratory.

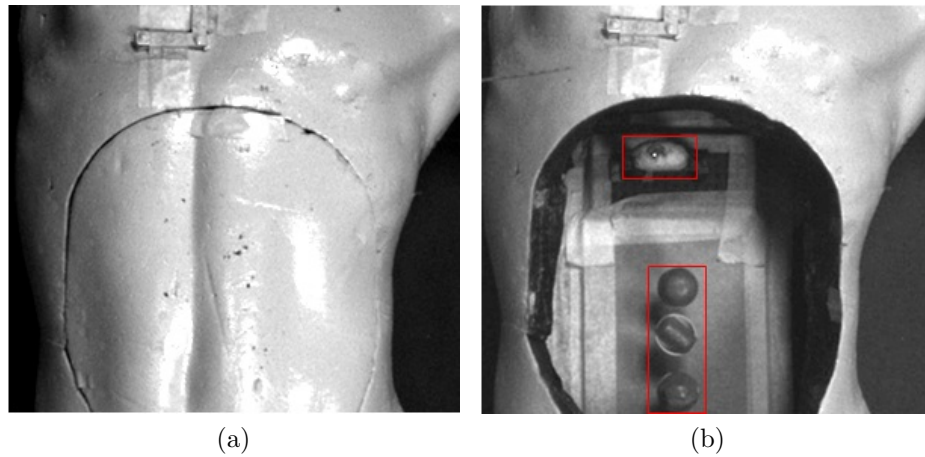


Figure 7-2. Mannequin with a removable lid and plasticine models inside. (a) With the lid in place for projection examination. (b) With the lid removed and plasticine models exposed for insertion verification.

An *in vivo* animal (porcine) study is conducted to evaluate AR display and registration. In order to create a phantom tumor in the live porcine liver, an

experimented radio-opaque gel substance was injected into the liver to mimic a tumor target for ablation. CT scan was carried out to locate the position of the tumor. The medical image data obtained was presented in a visualization model on the planning interface. Segmentation of the tumor and periphery critical anatomic structures were then carried out for the subsequent 3D model based preoperative surgical planning. Figure 7-3 shows the preparation for the CT scan and the resultant surgical plan based on the porcine 3D model which was reconstructed from the CT data set.

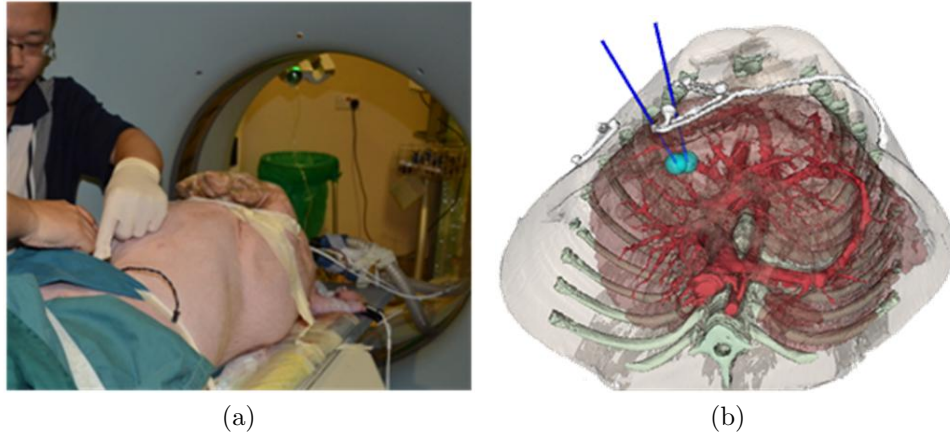


Figure 7-3. Deploying markers on the porcine surface before CT scanning (a) and surgical planning based on porcine anatomy model (b).

7.1 Projection Accuracy Evaluation

Pixel correspondence between the projector and the cameras of the stereo vision device was established by calibration of the ProCam subsystem. Projection accuracy was evaluated by projecting a chessboard pattern on the mannequin (Figure 7-2). The mannequin body was made of polystyrene with a 1.5 mm painted layer. The surface of the mannequin body was uneven and rough (Figure 7-2a).

Figure 7-4 compares the distorted projection and its corrected result on the target region of the mannequin body. The projection, which was distorted by

mannequin's uneven surface in Figure 7-4a has been corrected in Figure 7-4b. Quantitative accuracy evaluation on this correction was examined by comparing the source image of the chess-board pattern with its corresponding corrected projection in the Figure 7-4b. Geometric errors were calculated on each corresponding square unit along the x and the y axis. Average distortion error was 0.68 mm and 0.73 mm respectively.

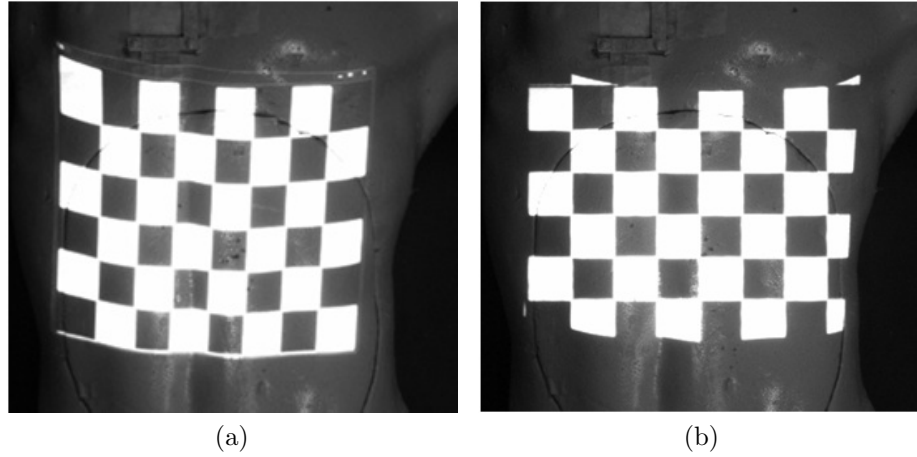


Figure 7-4. Projection ((a) distorted (b) corrected) on the mannequin.

Table 7.1: Deviation statistics for projection correction

| | |
|---|----------|
| Mean deviation from original points | 0.253 mm |
| Greatest deviation from original points | 0.627 mm |
| Smallest deviation from original points | 0.247 mm |
| Median deviation from original points | 0.280 mm |
| Standard deviation of deviation | 0.131 mm |
| Root Mean Square Error (RMSE) | 0.260 mm |

By projecting a group of feature points on a white blank paper, the varying motion centers and the corresponding feature points are verified when the paper is smoothly folded and unfolded. Figure 7-5 (top) shows motion of the corresponding feature points during the paper unfolding procedure. Their motion centers are detected and the feature correspondences are updated. Display of the chessboard pattern in Figure 7-5 (bottom) demonstrates a new projector-camera mapping

from the dynamic projector-camera system calibration is achieved and it can be used for pre-correcting the projector image for projection on a dynamic arbitrary surface. Although the projection of checkerboard pattern can be dynamically recovered, projection defects may appeared on the sharp corners when the paper is unsmoothly folded. This is because some feature correspondences are missed during the irregular changes. By comparing the chessboard square units on the projection images with the standard ones, we can investigate variation of projection errors during the paper motion cycle. The error is represented by percentage of the distorted areas of chessboard units. The average error of this chessboard projection during the motion cycle is 4% to 11 %. This manual-operated dynamic projection surface is only used to test ProCam calibration on an arbitrary dynamic surface. However, the motion of the surface may not satisfy the assumption that motion is "continuous and smooth" and "surface deforms as little as possible between small time intervals" mentioned in Chapter 3. Therefore, this result has large calibration errors. The experiment results in Section 7.3 (Figure 7-8) demonstrate the ProCam system calibration and image correction on a *vivo* swine's skin can satisfy the surgical accuracy requirement (Table 7.2)).

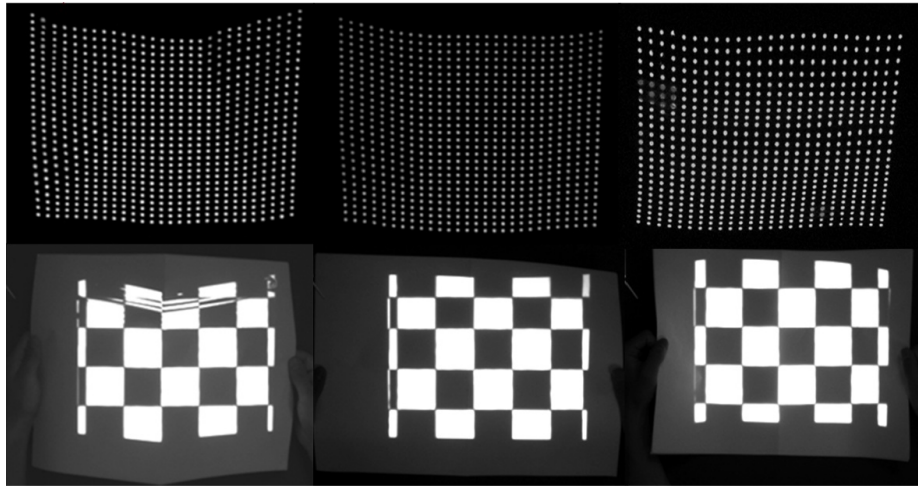


Figure 7-5. Projection of a checkerboard pattern on a dynamic blank paper.

7.2 Registration Evaluation

The four implanted plasticine models as well as the fiducial marker frame attached to the mannequin body were scanned by a CT machine before registration. With structured light illumination on the mannequin body, its surface data acquired by ProCam system was used to register patient space to model space using the surface matching algorithm. Evaluation of this registration result was designed by projecting the registered virtual plasticine models on the mannequin body to examine their projection with respect to the location of the real models. In order to exclude the effect of geometric distortion on registration error, the projection of the image with the plasticine models that was captured from the real camera's view was used as reference (Figure 7-6).

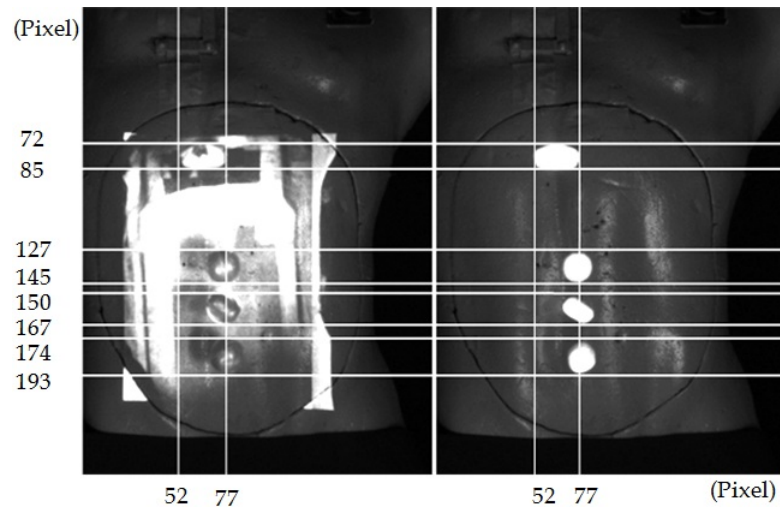


Figure 7-6. Examination of model-patient registration by overlaying the plasticine models on the real ones which were placed inside the mannequin. Projection of the image with the real plasticine models captured from the real camera's view (left) was considered as projection with expected position. Projection of the registered virtual plasticine models captured from the virtual camera's view in the anatomic model space (right) was tested.

The projection in Figure 7-6(left) shows the actual plasticine models' positions inside the mannequin with projective geometric corrections. The registration error can be determined by finding the difference between the image of the registered

virtual plasticine models projected on the mannequin in Figure 7-7 (right), which was also subjected to the geometric correction, and the projected image of the real models in Figure 7-6(left).

The boundary positions of the plasticine models in Figure 7-6 were measured and compared. Maximum and minimum errors along the contours of the plasticine models were calculated, to determine the positive maximum, negative maximum, positive minimum and negative minimum errors (Figure 7-7) for each plasticine model in Figure 7-6. After eliminating the model reconstruction error ranging from 0.26 mm to 0.55 mm, the error range of model-patient registration was then derived. Figure 7-7 shows the registration error distribution of the projected plasticine models.

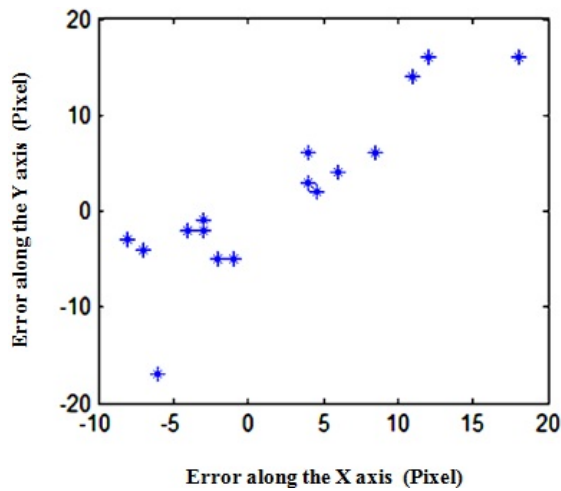


Figure 7-7. Registration errors of the four plasticine models.

Average errors for the four models were 8.7 pixels and -4.5 pixels along the x axis, and 9.8 pixel and -6.0 pixels along the y axis. According to the resolution and the sensor type of the cameras (780×580 and $1/2''$ CCD), the average error was 0.62 mm and -0.58 mm along the x axis and 0.66 mm and -0.59 mm along the y axis. The maximum error was 0.69 mm and the minimum error 0.27 mm. Thus, according the accuracy standard of phantom image registration suggested

by (Hostettler et al., 2010), the registration accuracy achieved in this study can satisfy the accuracy requirements in clinical practice.

7.3 Evaluation of Augmented Interaction

With this direct AR interface, the surgeon can inspect the preoperative surgical planning directly on the current patient body. The position of the preplanned ablation model and the anatomic structure, especially of critical organs and vessels can be examined by the surgeon. Surgical risk of the execution of the preoperative plan in the real surgical environment can be estimated. Similarly, the preplanned surgical data with respect to the practical robot setup can be examined. With regard to the inappropriate planning data such as inappropriate insertion ports and trajectories, surgeons can interactively revise the corresponding preplanned data on the computer workstation and view the revised plan on the AR interface immediately. In the *vivo* experiment (Figure 7.8) that was conducted in National University Hospital, we tested the ProCam in a real surgical environment (operation room). A surgeon tested this ProCam-based surgical AR guidance system and examined the AR display for surgical planning and guidance. He validated its feasibility in the clinical practice. Figure 7-8 shows an intraoperative supervisory guidance based on the surgical planning.

The RF needles placement under the guidance of the projector-based AR was examined (Figure 7-9 and Figure 7-10). In this mannequin study, the first plasticine model (Figure 7-2b) was used to represent a tumor in the liver of the mannequin body. In order to simulate the immersive anatomical structures around a liver tumor, an open source medical data comprising liver, vessels and ribs was used in the experiment. The tumor's position was manually defined within the liver volume. With the above registration results, the anatomical structure

models as well as the planning data were projected on the mannequin body.

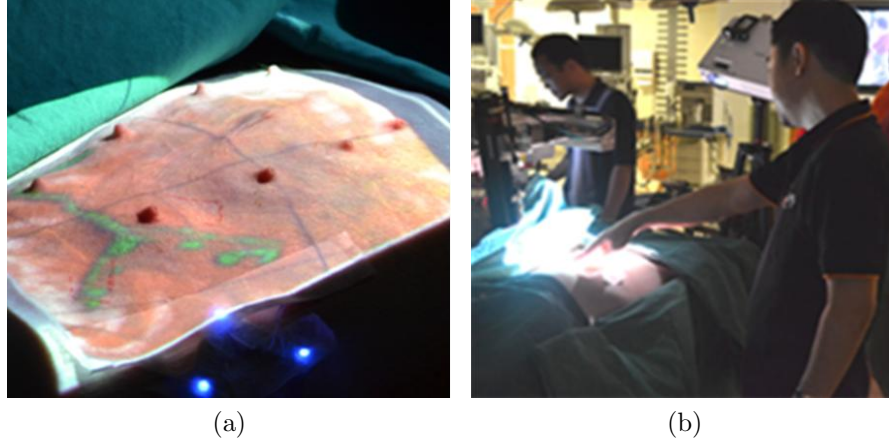


Figure 7-8. Spatial AR based visual guidance. (a) AR display of planning data on the porcine belly. (b) Surgeons can provide their feedback based on the AR interface.

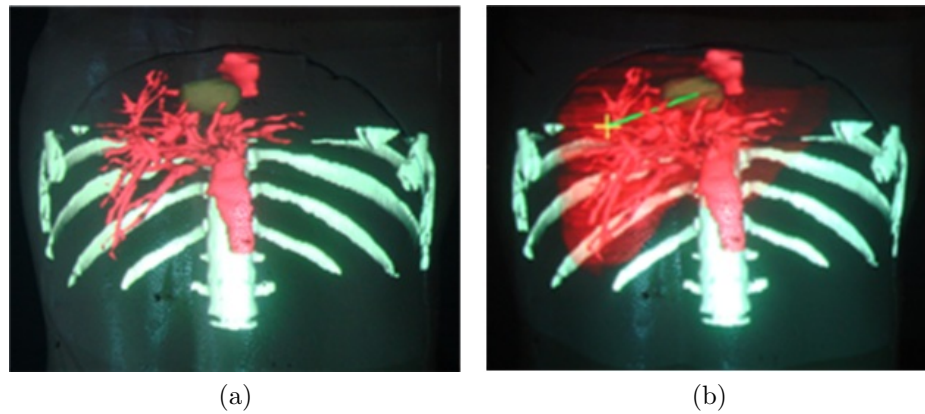


Figure 7-9. ProCam-based surgical AR display on the mannequin body. (a) AR display of the critical structures, vessels and tumor. (b) The preplanned insertion point and the insertion trajectory were highlighted on the patient surface for the first needle implant.

Figure 7-9 shows an overview of the AR environment with the anatomical models overlaying on the mannequin's surface. By changing the anatomical models' properties (e.g. transparency and orientation), specific structures could be clearly detected (Figure 7-9a). This facilitates the surgeon in acquiring knowledge of needle's position with respect to different anatomical structures. It also assists the surgeon in comparing the preoperative planning with the current patient position. When needle insertion begins, the preplanned insertion point and insertion trajec-

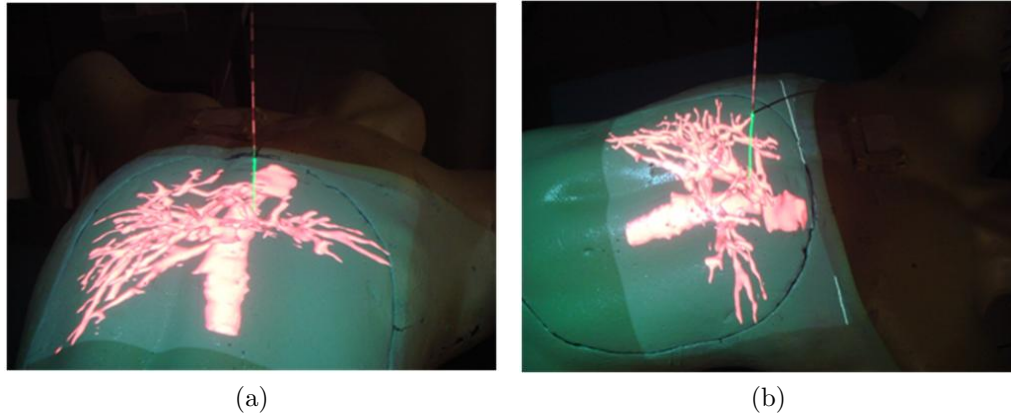


Figure 7-10. Viewer's position dependent AR display of the needle path for augmented interaction between the real and virtual needle segments.

tory are highlighted on the mannequin's surface to guide the RF needle's implant (Figure 7-9b).

Figure 7-10 demonstrates that viewer's position dependent AR display of the virtual needle segment inside the mannequin body with depth information can be achieved. Tracking viewer's position and incorporating this information in the registration process enables the viewer to comprehend the depth information so as to overcome the lack of depth perception typical in projection images. However, a limitation that may rise from the viewer-dependent display is visual conflicts between display of the rotated surgical model and actual position of the patient body. Figure 7-11 illustrates the augmented interaction procedure for needle insertion. The needle's actual trajectory due to the surgical was displayed in real time. Based on the overlapping ablation model constructed in preoperative planning, three insertions were required to completely ablate the entire tumor model (Figure 7-11c).

The actual insertion trajectory was evaluated by comparing it with its corresponding trajectory from the preplanned model. By splitting the tumor model (Figure 7-12), actual ablation point and its insertion trajectory could be acquired.

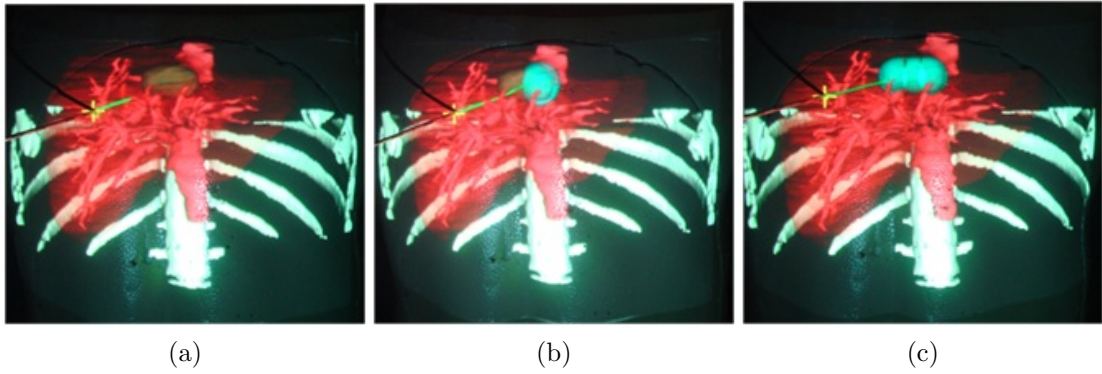


Figure 7-11. Augmented needle insertion process: (a)-(b) direct augmented interaction with the RF needle insertion providing surgeon's visual feedback for supervision of robotic execution; (c) insertion completed with the overlapping ablation model.

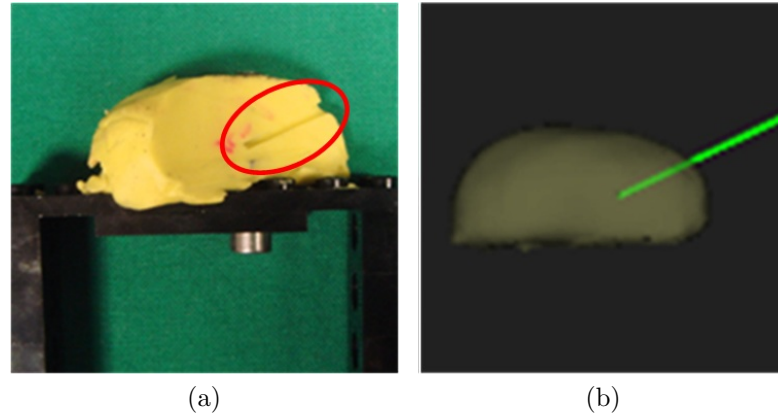


Figure 7-12. Comparison of the actual trajectory of the RF needle insertion (a) with its pre-planned one generated in the preoperative planning (b).

The vector for actual insertion trajectory could be reconstructed with the insertion point on the surface of the tumor model and the ablation point within the model. The error of the needle placement along the actual insertion trajectory with respect to the preplanned one was then determined.

As listed in Table 7.2, 36 insertions for 12 groups were tested on the tumor model, the phantom organ 1 (Figure 7-13). The mean error for the ablation points within the model was 1.82 mm. Mean deviation of these implants was 1.34 mm. The less than 2 mm error is acceptable clinically. Experiments were also conducted on the other three plasticine models inside the mannequin body

(phantom organs 2, 3, 4 in Table 7.2). Taking into consideration the accuracy of the robotic implementation with mean error of 0.53 mm, the mean error of projection of the virtual needle was 1.86 mm.

Table 7.2: Experimental data for robot-assisted needle insertion.

| No. of the Target | No. of Insertion | Mean Error | Mean Deviation |
|-------------------|------------------|------------|----------------|
| phantom organ 1 | 36 | 1.82 mm | 1.34 mm |
| phantom organ 2 | 30 | 1.88 mm | 1.92 mm |
| phantom organ 3 | 28 | 1.94 mm | 2.17 mm |
| phantom organ 4 | 30 | 1.79 mm | 1.84 mm |

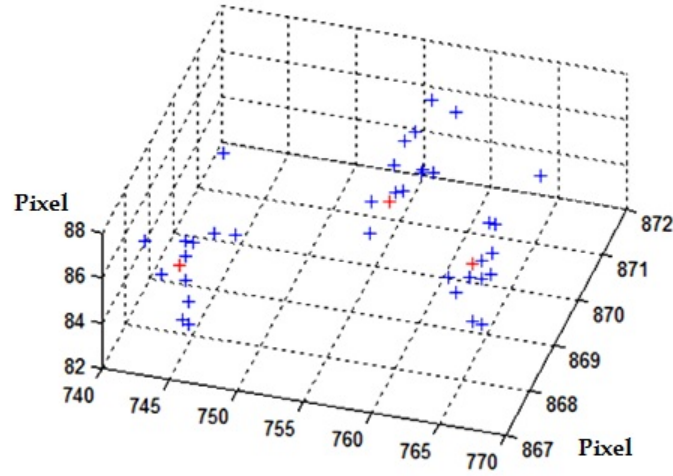


Figure 7-13. Needle implants for the tumor model test. The red crosses represent the expected needle placements.

Since the experiments for accuracy evaluation were performed on a static mannequin body, possible errors due to patient breathing was not considered during the intraoperative needle insertion procedures. The preplanned data should adapt to the changed anatomy during surgery. This may be achieved by intraoperative CT imaging and simulation of organ motion with preoperative images (Rieder et al.,

2011). The simulation of organ motion with preoperative 3D CT or MRI images may predict the tumor position during intraoperative intervention. According to Hostettler et al.'s study (Hostettler et al., 2010), accuracy of the estimated tumor position could be within 23 mm at 50 Hz of patient free respiration which meets the clinic requirement.

In order to resolve the dynamic projection problems such as projection correction and registration due to patient movement or breathing, structured-light illumination and hybrid method (EKF and bending energy) in Chapter 3 can be used to estimate the ProCam pixel correspondence. The updated pixel correspondence can be established by real-time tracking of the motion field of the projection on the patient surface.

Recently, Kinect system draws increasing interest in the field of vision-based tracking and recognition. It has advantages in acquiring the depth information and surgeon-AR interaction. However, its frame rate (9 - 30 f/p) and accuracy (4.5 mm - 4 cm) may not meet the surgical or medical requirements as CCD camera does. According to an academic analysis of Kinect system (Khoshelham and Elberink, 2012), their experimental results demonstrate that the random error of depth measurement increases with increasing distance to the sensor, and ranges from a few millimeters up to about 4 cm at the maximum range of the sensor. The quality of the data is also found to be influenced by the low resolution of the depth measurements. Therefore, the Kinect may not be suitable in surgical applications.

7.4 Parallel Acceleration with GPU

In the hybrid algorithm for system calibration, three computationally expensive phases are involved for each frame including recursive EKF, local curvature com-

putation and minimization of bending energy (Chapter 3). Massive image data processing is also a time consuming task. In order to speeding the implementation of the data processing and matrix operation on the massive feature correspondence, GPU-based paralleled computation is developed to enhance the calculation efficiency in the computationally expensive process.

The Algorithm 2 describes the algorithm performed on each frame to obtain the edge map using CUDA. The structure of the blocks and threads in CUDA illustrated in (Olmedo et al., 2012) demonstrates that the location of a single thread in the complete matrix is a combination of a block number and a thread number in the block. The index i is one pixel in position i and the indexes $i \times 3, i \times 3 + 1$ and $i \times 3 + 2$ refer to the red, green and blue channels respectively. Canny edge detector is applied on each channel and then a logic OR operation is implemented.

Algorithm 2 CUDA-based edge map generation

1. /*I: input image (current frame)*/
 2. /*C: operation by Canny detector*/
 3. /*EGM: edge map*/
 4. For each GPU processing i ,
 5. $Proc(i) = blockIDx.x \times (blockDim.x \times blockDim.y)$
 6. $+blockDim.x \times threadIDx.y \times threadIDx.x$
 7. $I[i \times 3] = C([i \times 3])$
 8. $I[i \times 3] = C([i \times 3])$
 9. $I[i \times 3] = C([i \times 3])$
 10. $EGM(I) = I[i \times 3] \text{ OR } I[i \times 3 + 1] \text{ OR } I[i \times 3 + 2]$
 11. endfor
-

In computation of EFK and bending energy minimization, the massive recursive matrix operation can be implemented concurrently among different iterations. Different blocks and block threads are defined as computational units for different basic matrix operation. These operations include matrix addition, matrix subtraction, matrix multiplication, transpose for matrix transposition, and `inv_spd` for inversion of a semipositive definite (SPD) matrix. For matrix inversion, the Cholesky factoring (Olmedo et al., 2012) can be used for computing semi-positive definite matrices. It is faster than the common LU factoring or Gaussian elimination used for general square matrices. During the iterative processing of EKF and bending energy, blocks and threads (Figure 7-14) are filled with the matrix operation from different iterations, which enables parallel computation of EFK and bending energy minimization simultaneously.

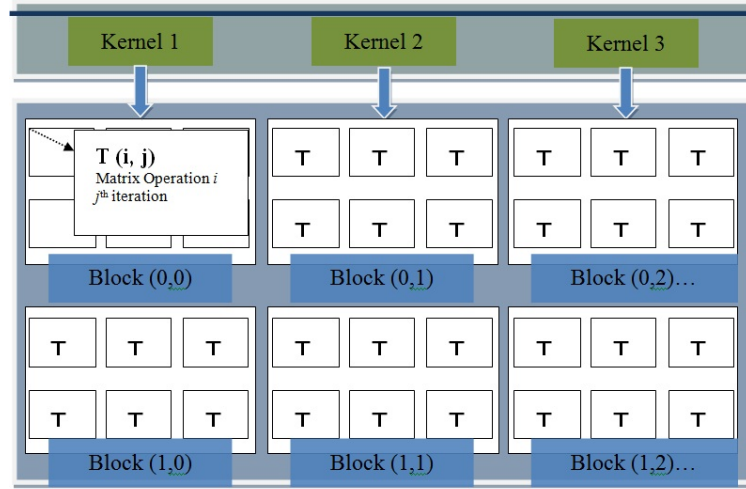


Figure 7-14. Parallel matrix operation based on CUDA structure for accelerating EFK tracking and bending energy minimization.

Figure 7-15a shows how the computation time increases as the massive matrix operation are implemented on CUDA structure and CPU. With matrix dimension and number of operating matrices increase, the CUDA-based parallel computing shows impressive advantages for this algorithm. This massive parallel computing ability performs an important role for the dynamical calibration including EKF

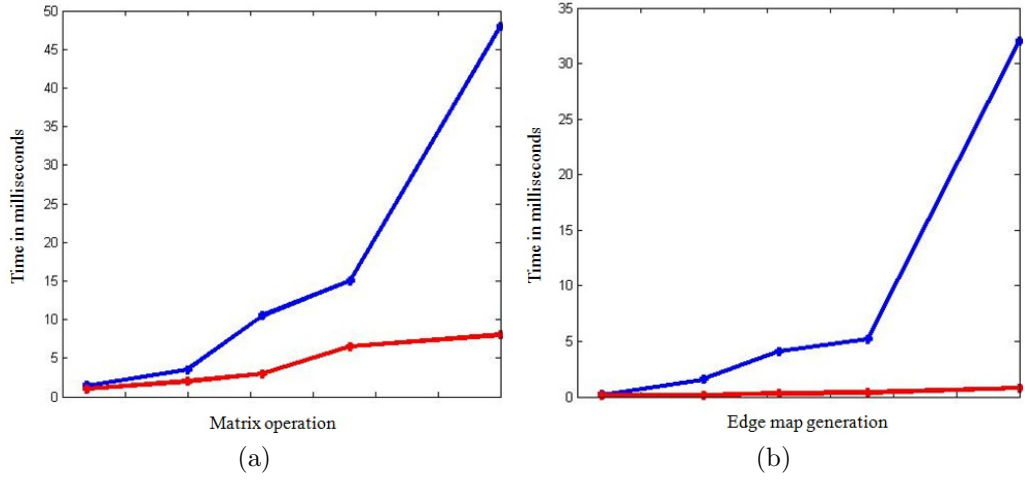


Figure 7-15. Performance comparison: (a) comparative graphic of matrix operation in the process of EFK computation and bending energy minimization with CUDA vs. CPU; (b) comparative graphic of edge-map generation with CUDA vs. CPU.

recursive tracking and curvature parameters computation. As can be seen in the Figure 7-15a, the computation time is approximately linear after more matrix operations are involved, while the computation time is significantly reduced with CUDA-based GPU. Figure 7-15b shows the performance of CUDA-based parallel computing on massive image data processing.

7.5 Summary

Our experiments and results are summarized and discussed in Chapter 7. We experiment projection-based AR display for visual feedback and navigation of needle insertion procedure in RF ablation treatment. Errors resulted from the projection correction and registrations are discussed. The relevant findings enable us to have a better understanding on projector-camera calibration, surface-based model-patient registration, augmented interaction for projection-based AR environment and their integration with robotic needle insertion. Besides, limitations on augmented needle insertion with a rigid needle, patient' breathing effects and integration of Kinect system as a guidance are discussed. These experimental

results and discussion are helpful for us to set up the roadmap of future work.

Chapter 8

CONCLUSION

This research explored a direct AR technology with ProCam system for surgical guidance. The accomplished work includes firstly geometric and radiometric correction methodology for ProCam-based AR display; secondly, registration among surgical model, patient body and projection; and thirdly the augmented interaction for the projection-based AR environment.

The underlying mechanisms of AR display and its registration were reviewed in Chapter 2. The image-based surgical guidance and interaction in current IGS were also discussed in the chapter. Our improved ProCam calibration method was described in Chapter 3. Our major contributions include pixel correspondence between the projector and camera with a phase-shift structured illumination, and a hybrid algorithm to minimize energy based curvature for dynamic pixel mapping. In Chapter 4, we presented our work on projective correction on irregular surfaces, which includes the principle and method of piecewise pre-warping, radiometric correction model as well as texture mapping for pixel value correction. Chapter 5 focuses on the registration of model projection with the corresponding organ ar-

on the patient's surface. We proposed two algorithms, surface matching and point-based registration for model-patient and patient-projection registration respectively. In Chapter 6, we developed a means of hand gesture recognition for direct HCI in ProCam-based surgical AR environment. Augmented needle insertion was proposed and developed for surgical supervisory guidance. All experiments and results were further discussed in details in Chapter 7 with a recapitulation of the issues of AR display, registration, augmented interaction as well as their integration. Evaluations of *in vivo* and *ex vivo* experiments were qualitatively and quantitatively reported in Chapter 7.

8.1 Summary of Contributions

This research establishes a navigational procedure for percutaneous surgical augmented intervention integrating a ProCam-based AR guidance and robotic implementation for augmented needle insertion. It provides surgeons a new way to supervise the robot-assisted image-guided surgery, which overcomes the visual and operational limitations in minimally invasive procedure. The contributions of this research are summarized as follows:

Proposed and designed an integrated ProCam system (Figure 7.4) for construction of surgical AR environment. Compared to the current ProCam system, this system consists of the motor-controlled ProCam modules, tracking devices and surgical robotic system, which are optimally deployed on a rectangular robot. The integrated system can not only display the projection-based AR information on a patient body, but also can use this AR information to guide surgeons' operation or robotic implementation.

Proposed a dynamic ProCam calibration method with EKF and bending energy

optimization (Chapter 3). This method provides a solution to find the pixel correspondence between the projector and camera from a dynamic object surface. In particular, there is no embedded pattern projection involved in this method.

Improved the accuracy of geometric correction with the pre-warping principle and texture mapping based pixel value correction. In current monitor-based and optical see-through AR display (Table 1.1), the computer-generated models are laid over the captured scene, thus their AR displays are closed and less immersive compared with the direct projection-based AR. However, the projection-based AR display is susceptible to irregular geometry and textures of the target surface. In this research, a texture mapping based pixel value correction is developed to directly map the pixel values in the normalized captured images to their corresponding pixels in the projector images. The method enables projection of surgical model to display geometrically correct on the patient (skin) surface.

Efficient registration algorithm and method were proposed and developed for ProCam-based surgical AR guidance system, among surgical models, patient body and model projection images. AR registration defines the spatial relationship between the virtual and real scenes, which is a common challenge in an AR construction. The registration work is particularly challenging for ProCam-based AR because it has to directly overlay the virtual models on the real objects rather than overlapping their images on a computer screen. For surgical AR guidance, it is also difficult to precisely relate the virtual models with its projection on the patient's (skin) surface. In addition, there are few reports on ProCam-based registration and augmented interactions for surgical guidance. In this research, a surface matching algorithm is developed for registration between the surgical model and the patient body while a point-based registration method for the projection and patient surface. This hybrid registration method can significantly reduce the num-

ber of deployed markers on the patient's body. This has enhanced the registration accuracy and efficiency, and minimized the chances that the surgical operation being obstructed by the markers.

Develop a hand gesture based direct user-AR interaction for the surgical AR environment. This non-contact interaction avoids sterilization problems and it is efficient for surgeons to provide their intraoperative feedback. Besides the user-AR interaction, mechanism of augmented needle insertion with surgical robot assistance is explored and tested. The impact of this augmented needle insertion on the clinical application can be immense and this may be one of the first inroads in this direction.

8.2 Future Work

There are many pending issues that have to be addressed for a successful ProCam-based AR system for surgical guidance. For example, in terms of augmented interaction, we are faced with the dilemma of dynamic shadow elimination which may be caused by users and surgical tools. The discrepancy between surgical model simulation and uncertainty of interaction between the surgical tools and anatomy is another problem (Maio and Salcudean, 2005). It is a topic worthy of investigation to accurately deliver intraoperative surgical information to surgeons.

We will continue to enhance immersive AR display with respect to efficiency, reliability and reproducibility, which may determine the applicability of ProCam system for clinical use. This AR guidance for clinical applications will be further evaluated in collaboration with National University Hospital (NUH, Singapore). The following specific topics are worthy of further investigation:

Dynamic Augmented Interaction: Since the experiments for the accuracy eval-

uation of augmented interaction were performed on a static mannequin body, the error caused by patient's breathing movements was not considered during the intraoperative needle insertion procedures. In clinical practice, the preplanned data need to be adapted to the dynamic anatomy during the surgery. Currently, intraoperative CT imaging and simulation of organ motion with preoperative images may provide acceptable solutions to this problem (Krempien et al., 2008). Simulation of the organ motion with preoperative 3D CT or MRI images (Hostettler et al., 2010, Wen et al., 2010a) could predict the tumor position in the intraoperative intervention. The reported accuracy for the simulation of the abdominal viscera positions, such as the liver, kidneys or spleen, at 50 Hz was within 23 mm.

Real-time Finite Element (FE) Simulation: The RF needle in this experiment was assumed to be a rigid body without bending from the insertion points to the final position. The RF needle may suffer bending, especially interaction with the soft inhomogeneous tissue. Recent studies (Chui et al., 2006) on FE model simulation of needle-based model suggested a method to reconstruct the needle model interaction with the tissue model in real time during insertion. Moreover, deformation occurring on the anatomic models could be also derived from this insertion simulation.

Bibliography

- S. Ait-Aoudia, F. Z. Benhamida, and M. A. Yousfi. Lossless compression of volumetric medical data. In *Proceedings of the 21th International Symposium on Computer and Information Sciences*, pages 563–571, Nov. 2006.
- T. Akinbiyi, C. E. Reiley, S. Saha, D. Burschka, C. J. Hasser, D. D. Yuh, and A. M. Okamura. Dynamic augmented reality for sensory substitution in robot-assisted surgical systems. In *Proceedings of IEEE Engineering in Medicine & Biology Society*, pages 567–570, Jan. 2006.
- K. M. Al-Aidaroos, A. Abu Bakar, and Z. Othman. Naive Bayes variants in classification learning. In *Proceedings of International Conference on Information Retrieval and Knowledge Management: Exploring the Invisible World*, pages 276–281, Jun. 2010.
- D. F. Amanatullah, M. T. Burrus, S. S. Sathappan, B. Levine, and P. E. D. Cesare. The application of minimally invasive surgical techniques. part I: Total hip arthroplasty. *Am J Orthop*, 41(10):e134–e139, Oct. 2012.
- J. B. Antoine Maintz and M. A. Vierger. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–37, Sep. 1998.
- K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-D

- point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(5):698–700, Sep. 1987.
- S. Audet and M. Okutomi. A user-friendly method to geometrically calibrate projector-camera systems. In *Proceedings of the IEEE International Workshop on Projector-Camera Systems*, pages 47–54, Jun. 2009.
- C. Baegert, C. Villard, P. Schreck, and L. Soler. Multi-criteria trajectory planning for hepatic radiofrequency ablation. *Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention(MICCA)*, 177: 676–684, Oct. 2007.
- M. Bajura and U. Neumann. Dynamic registration correction in video-based augmented reality systems. *Computer Graphics and Applications*, 15(5):52–60, Jan. 2005.
- H. Benko, R. Jota, and A. Wilson. Miragetable: Freehand interaction on a projected augmented reality tabletop. In *Proceedings of Human Factors in Computing Systems (CHI)*, pages 199–208, May 2012.
- O. Bimber, A. Emmerling, and T. Klemmer. Embedded entertainment with smart projectors. *Computer*, 38(1):48–55, Jan. 2005a.
- O. Bimber and R. Raskar. *Spatial Augmented Reality - Merging Real and Virtual Worlds*. A K Peters, 1st edition, 2005.
- O. Bimber, G. Wetzstein, A. Emmerling, and C. Nitschke. Enabling view-dependent stereoscopic projection in real environments. In *IEEE/ACM International Symposium on Mixed and Augmented Reality(ISMAR05)*, pages 14–23, Sep. 2005b.

- J. Y. Bouguet. Camera calibration toolbox for Matlab. *Available at [http :
//www.vision.caltech.edu/bouguetj/calibdoc/index.html](http://www.vision.caltech.edu/bouguetj/calibdoc/index.html)*, Jul. 2010.
- G. Bradski and A. Kaehler. *Learning OpenCV: Computer Vision with the OpenCV Library*. O'Reilly Media, 1st edition, 2008.
- C. Brauer-Burchardt, A. Breitbarth, P. Kuhmstedt, I. Schmidt, M. Heinze, and G. Notni. Fringe projection based high-speed 3D sensor for real-time measurements. In *Proceedings of Optical Measurement Systems for Industrial Inspection VII*, pages 192–197, May 2011.
- M. Burns. An insider’s look at ford’s virtual reality design tools. *Available at [http :
//techcrunch.com/2010/05/22/an – insiders – look – at – fords – virtual –
reality – design – tools/](http://techcrunch.com/2010/05/22/an-insiders-look-at-fords-virtual-reality-design-tools/)*, 2010.
- N. D. Cahill, C. M. Williams, S. Chen, L. A. Ray, and M. M. Goodgame. Incorporating spatial information into entropy estimates to improve multimodal image registration. In *Proceedings of IEEE International Symposium on Biomedical Imaging: Nano to Macro*, pages 832–835, Apr. 2006.
- J. Chaoui, C. Hamitouche, E. Stindel, and C. Roux. Recognition-based segmentation and registration method for image guided shoulder surgery. In *International Conference on IEEE Engineering in Medicine and Biology Society*, pages 6212–6215, Aug. 2011.
- D. Chekhlov, M. Pupilli, W. M. Cuevas, and A. Calway. Robust real-time visual slam using scale prediction and exemplar based feature description. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pages 1–7, Jun. 2007.
- M. H. Chen, K. Yan, M. W. Zou, L. Solbiati, J. B. Liu, and Y. Dai. Large liver tumors: protocol for radiofrequency ablation and its clinical application

- in 110 patients—mathematic model, overlapping mode, and electrode placement process. *Radiology*, 232(5):206–271, Sep. 2004.
- X. Chen, J. Xi, Y. Jin, and J. Sun. Accurate calibration for a camera-projector measurement system based on structured light projection. *Optics and Lasers in Engineering*, 47(34):310–319, Apr. 2009.
- L. Chmielewski and D. Kozinska. Image registration. In *Proceedings of Conference on Computer Pattern Recognition Systems*, pages 26–29, May 2003.
- C.-K. Chui, E. Kobayashi, X. Chen, T. Hisada, and I. Sakuma. Transversely isotropic properties of porcine liver tissue: experiments and constitutive modelling. *Medical and Biological Engineering and Computing*, 45(1):99–106, Jan. 2007.
- C.-K. Chui, E. Kobayashi, T. Hisada, and I. Sakuma. Combined compression and elongation experiments and non-linear modelling of liver tissue for surgical simulation. *Medical and Biological Engineering and Computing*, 42(6):787–798, Nov. 2004.
- C.-K. Chui, S. H. Teoh, C. J. Ong, J. H. Anderson, and I. Sakuma. Integrative modeling of liver organ for simulation of flexible needle insertion. In *Proceedings of 9th International Conference of Control, Autom., Robot. Vision (ICARCV)*, pages 1–6, Aug. 2006.
- D. Cotting, M. Naef, M. Gross, and H. Fuchs. Embedding imperceptible patterns into projected images for simultaneous acquisition and display. In *Processing of IEEE International Symposium on Mixed and Augmented Reality (ISMAR04)*, pages 100–109, Jun. 2004.
- D. Cotting, R. Ziegler, M. Gross, and H. Fuchs. Adaptive instant displays: Contin-

- uously calibrated projections using per-pixel light control. *Computer Graphics Forum*, 24(3):705–71, Apr. 2005.
- A. Davison, W. Mayol, and D. Murray. Real-time localization and mapping with wearable active vision. In *Proceedings of IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 18–27, Oct. 2003.
- A. Davison, I. Reid, N. D. Molton, and O. Stasse. Monoslam: Real-time single camera slam. *IEEE Transactions on Pattern Anal. Mach. Intell.*, 29(6):1052–1067, Jun. 2007.
- A. de Landgraaf. Interaction between users and augmented reality systems: Human-computer interaction of the future. *Human-Computer Interaction*, 11(2):125–156, Nov. 2009.
- G. D. Dodd, M. S. Frank, M. Aribandi, S. Chopra, and K. N. Chintapalli. Radiofrequency thermal ablation: computer analysis of the size of the thermal injury created by overlapping ablations. *AJR Am J Roentgenol*, 177:777–782, Oct. 2001.
- J. Drarni, P.F. Sturm, and S. Roy. Methods for geometrical video projector calibration. *Machine Vision and Applications Journal*, 23(1):79–89, Apr. 2012.
- Q. Du and X. Zhang. Miragetable: Freehand interaction on a projected augmented reality tabletop. In *Proceedings of IEEE International Conference on Industrial Technology (ICIT)*, pages 1–6, May 2008.
- G. Falcao, N. Hurtos, J. Massich, and D. Fofi. Projectorcamera calibration toolbox. Available at [http : //code.google.com/p/procamcalib](http://code.google.com/p/procamcalib), 2009.
- O. D. Faugeras and G. Toscani. The calibration problem for stereo. *Proceedings of the IEEE Computer Vision and Pattern Recognition*, pages 15–20, Jun. 1986.

- G. Fichtinger, A. Deguet, K. Masamune, E. Balogh, G. S. Fischer, H. Mathieu, R. H. Taylor, S. J. Zinreich, and L. M. Fayad. Image overlay guidance for needle insertion in ct scanner. *IEEE Transactions on Biomedical Engineering*, 52(8): 1415–1424, Aug. 2005.
- J. Frund, J. Gausemeier, C. Matysczok, and R. Radkowski. Using augmented reality technology to support automobile development. *Lecture Notes in Computer Science*, 31(68):289–298, 2005.
- M. Fujigaki and Y. Morimoto. Shape measurement with grating projection using whole-space tabulation method. *Journal of of Structural Engineering and Mechanics*, 8(4):92–98, Aug. 2008.
- K. Fujii, M. D. Grossberg, and S. K. Nayar. projector-camera system with real-time photometric adaptation for dynamic environments. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2005)*, pages 814–821, Jun. 2005.
- J. Gao, A. Kosaka, and A. C. Kak. A multi-kalman filtering approach for video tracking of human-delineated objects in cluttered environments. *Computer Vision and Image Understanding*, 102(3):260–316, Jun. 2006.
- S. Garrean, J. Hering, A. Saied, W. S. Helton, and N. J. Espat. Radiofrequency ablation of primary and metastatic liver tumors: a critical review of the literature. *The American Journal of Surgery*, 195(3):508–520, Apr. 2008.
- J. Garstka and P. Gabriele. View-dependent 3D projection using depth-image-based head tracking. In *Proceedings of IEEE International Workshop Projector-Camera Systems (PROCAMS)*, pages 176–181, Jun. 2011.
- K. A. Gavaghan, M. Peterhans, T. O. Santos, and S. Weber. A portable image

- overlay projection device for computer-aided open liver surgery. *IEEE Transactions on Biomedical Engineering*, 58(6):1855–1864, Dec. 2011.
- G. Geng. Structured-light 3d surface imaging: A tutorial. *Advances in Optics and Photonics*, 3(2):128–160, Dec. 2011.
- T. Guan and C. Wang. Registration based on scene recognition and natural features tracking techniques for wide-area augmented reality systems. *IEEE Transactions on Multimedias*, 11(8):1393–1406, Nov. 2009.
- M. Harders, G. Bianchi, B. Knoerlein, and G. Szekely. Calibration, registration, and synchronization for high precision augmented reality haptics. *IEEE Transactions on Visualization and Computer Graphics*, 15(1):138–149, Jan. 2009.
- R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2003.
- M. M. Hasan and P. K. Mishra. Features fitting using multivariate gaussian distribution for hand gesture recognition. *International Journal of Computer Science & Emerging Technologies*, 3(2):76–85, Jul. 2010.
- W. A. Hoff, K. Nguyen, and T. Lyon. Computer vision-based registration techniques for augmented reality. In *Proceedings of the Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments (SPIE1996)*, pages 18–22, Oct. 1996.
- A. Hostettler, S. A. Nicolau, Y. Remond, J. Marescaux, and L. Soler. A real-time predictive simulation of abdominal viscera positions during quiet free breathing. *Progress in Biophysics & Molecular Biology*, 103(2-3):169–184, May 2010.
- P. C. Hu, N. Li, and J. J. Zhou. Improved camera self-calibration method based on circular points. *Opto-Electronic Engineering*, 34(12):54–60, 2007.

- P. S. Huang, S. Zhang, and F. P. Chiang. Trapezoidal phase-shifting method for three-dimensional shape measurement. *Optical Engineering*, 44(12):123601–1–123601–8, Dec. 2005.
- H. Iwata. *Haptic interfaces*, chapter Haptic interfaces, pages 153–166. The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Application. M.E. Sharpe, 2003.
- T. Kahn and H. Busse. *Interventional Magnetic Resonance Imaging*. Springer, 2nd edition, 2012.
- F. Karray, M. Alemzadeh, J. Saleh, and M. Arab. Human-computer interaction: Overview on state of the art. *International Journal on Smart Sensing and Intelligent Systems*, 1(1):137–159, Mar. 2008.
- Z. Khan and A. Ibraheem. Hand gesture recognition: A literature review. *International Journal of Artificial Intelligence & Applications (IJAIA)*, 3(4):161–174, Jul. 2012.
- K. Khoshelham and S. O. Elberink. Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors*, 12(2):1437–1454, Dec. 2012.
- N. Kock. E-collaboration and e-commerce in virtual worlds: The potential of second life and world of warcraft. *International Journal of e-Collaboration*, 4(3):1–13, Sep. 2008.
- K. Konishia, M. Hashizumeb, M. Nakamotod, Y. Kakejib, I. Yoshinoc, A. Take-tomic, Y. Satod, S. Tamurad, and Y. Maehara. Augmented reality navigation system for endoscopic surgery based on three-dimensional ultrasound and computed tomography: Application to 20 clinical cases. *Computer Assisted Radiology and Surgery*, Jun. 2005.

- R. Krempien, H. Hoppe, L. Kahrs, S. Daeuber, O. Schorr, and G. Eggers. Projector-based augmented reality for intuitive intraoperative guidance in image-guided 3D interstitial brachytherapy. *International Journal of Radiation Oncology*Biolog*Physics*, 70(3):944–952, Jul. 2008.
- V. S. Kulkarni and S. D. Lokhande. Appearance based recognition of american sign language using gesture segmentation. *International Journal on Computer Science and Engineering (IJCSE)*, 2(3):560–565, Feb. 2010.
- K. N. Kutulakos. Calibration-free augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 4(1):1–20, Jan. 1998.
- D. Lanman and G. Taubin. Build your own 3D scanner: 3D photography for beginners. *Proceeding of SIGGRAPH '09 ACM SIGGRAPH 2009 Courses*, (8), 2009.
- B. Li and I. Sezan. Automatic keystone correction for smart projectors with embedded camera. In *Proceedings of International Conference on Image Processing*, pages 2829–2832, Oct. 2004.
- P. Li and J.N. Wang. Overview of camera calibration methods. *Shanxi Electronic Technology*, 4:77–79, 2007.
- H. Liao, T. Inomata, I. Sakuma, and T. Dohi. Three-dimensional augmented reality for MRI-guided surgery using integral videography auto stereoscopic-image overlay. *IEEE Transactions on Biomedical Engineering*, 57(6):1476–1486, Jul. 2010.
- J. Liu, B. C. Vemuri, and J. L. Marroquin. Local frequency representations for robust multimodal image registration. *IEEE Transactions on Medical Imaging*, 4(3):462–469, May 2002.

- N. Liu, B. C. Lovell, P. J. Kootsookos, and R. I. A. Davis. Model structure selection & training algorithms for an hmm gesture recognition system. In *Proceedings of International Workshop on Frontiers in Handwriting Recognition*, 2004.
- S. P. D. Maio and S. E. Salcudean. Interactive simulation of needle insertion models. *IEEE Transaction on Biomedical Engineering*, 52(7):1167–1179, Jul. 2005.
- A. Majumder and M. S. Brown. *Practical Multi- projector Display Design*. A. K. Peters, 1st edition, 2007.
- A. Majumder, D. Jones, M. McCrory, M. E. Papka, and R. Stevens. Using a camera to capture and correct spatial photometric variation in multi-projector displays. In *Proceedings of IEEE International Workshop on Projector-Camera Systems (ProCams)*, pages 1–8, Oct. 2003.
- S. Malik, G. Roth, and C. McDonald. Robust corner tracking for real-time augmented reality. In *Proceedings of International Conference on Vision Interface*, pages 399–406, May 2002.
- P. Manjusha and U. Bhosle. Registration of translated and rotated images using finite fourier transforms. *International Journal of Image Processing*, 5(3):245–703, Sep. 2011.
- D. Moreno and G. Taubin. Simple, accurate, and robust projector-camera calibration. In *Proceedings of 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, pages 464–471, Oct. 2012.
- N. Navab, J. Traub, T. Sielhorst, M. Feuerstein, and C. Bichlmeier. Action- and workflow-driven augmented reality for computer-aided medical procedures. *Computer Graphics and Applications*, 27(5):10–14, Oct. 2007.

- S. K. Nayar, H. Peri, M. D. Grossberg, and P. N. Belhumeur. A projection system with radiometric compensation for screen imperfections. In *Proceedings of IEEE International Workshop on Projector-Camera Systems (ProCams)*, Oct. 2003.
- A. Y. C. Nee, S. K. Ong, G. Chryssolouris, and D. Mourtzis. Augmented reality applications in design and manufacturing. *CIRP Annals - Manufacturing Technology*, 61(2):657–679, Nov. 2012.
- S. Nicolau, L. Soler, D. Mutter, and J. Marescaux. Augmented reality in laparoscopic surgical oncology. *Surgical Oncology*, 20(3):189–201, Sep. 2011.
- S. A. Nicolau, X. Pennec, L. Soler, X. Buy, A. Gangi, N. Ayache, and J. Marescaux. An augmented reality system for liver thermal ablation: design and evaluation on clinical cases. *Med Image Anal*, 13(3):494–506, Jun. 2009.
- E. Olmedo, J. Calleja, A. Benitez, and M. A. Medina. Point to point processing of digital images using parallel computing. *International Journal of Computer Science*, 9(3):251–276, May 2012.
- S. K. Ong, M. L. Yuan, and A. Y. C. Nee. Registration using projective reconstruction for augmented reality systems. In *Proceedings of ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry (VRCAI04)*, pages 286–291, Jan. 2004.
- H. Park, M.-H. Lee, B.-K. Seo, J.-I. Park, and M.-S. Jeong. Simultaneous geometric and radiometric adaptation to dynamic surfaces with a mobile projector-camera system. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(1):110–115, Apr. 2008.
- J. Park. Augmented reality based re-formable mock-up for design evaluation. In *Proceedings of the 2008 International Symposium on Ubiquitous Virtual Reality*, pages 17–20, Oct. 2004.

- J. Park and M. Kim. Interactive display of image details using a camera-coupled mobile projector. In *Proceedings of Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 9–16, Jun. 2010.
- A. Patriciu, M. Awad, S. Solomon, M. Choti, D. Mazilu, L. Kavoussi, and D. Stoianovici. Robotic assisted radio-frequency ablation of liver tumors—randomized patient study. *Med Image Comput Comput Assist Interv*, 8(Pt2): 526–533, Aug. 2005.
- J. Rekimoto and K. Nagao. The world through the computer: Computer augmented interaction with real world environments. In *Proceedings of Symposium on User Interface Software and Technology (UIST’95)*, pages 29–36, Nov. 1995.
- C. Rieder, T. Kroeger, C. Schumann, and H. K. Hahn. GPU-based real-time approximation of the ablation zone for radiofrequency ablation. *IEEE Transaction on Visualization and Computer Graphicss*, 17(12):1812–1821, Dec. 2011.
- F. Sadlo, T. Weyrich, R. Peiker, and M. Gross. A practical structured light acquisition system for point-based geometry and texture. In *Proceedings of Eurographics/IEEE VGTC Symposium*, pages 89–145, Jun. 2005.
- J. Salvi, X. Armangue, and J. Batlle. A comparative review of camera calibrating methods with accuracy evaluation. *Pattern Recognition*, 35(7):1617–1635, Jul. 2002.
- J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, Apr. 2004.
- K. Satoh, M. Anabuki, H. Yamamoto, and H. Tamura. A hybrid registration method for outdoor augmented reality. In *Proceedings of IEEE and ACM International Symposium on Augmented Reality*, pages 67–76, Jan. 2001.

- A. Seitel, M. Engel, C. M. Sommer, B.A. Radeleff, C. Essert-Villard, C. Baegert, M. Fangerau, K. H. Fritzsche, K. Yung, H.P. Meinzer, and L. Maier-Hein. Computer-assisted trajectory planning for percutaneous needle insertions. *Medical Physics*, 38(6):3246–3259, Oct. 2011.
- P. Shi, A. Sinusas, R. T. Constable, E. Ritman, and J. Duncan. Point-tracked quantitative analysis of left ventricular motion from 3d image sequences. *IEEE Transactions on Medical Imaging*, 19(1):36–50, Aug. 2000.
- T. Sielhorst, M. Feuerstein, and N. Navab. Advanced medical displays: A literature review of augmented reality. *Journal of Display Technology*, 4(4):451–467, Jul. 2008.
- L. Soler, S. Nicolau, J. Schmid, C. Koehl, J. Marescaux, X. Pennec, and N. Ayache. Virtual reality and augmented reality in digestive surgery. In *Proceedings of the 3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 278–285, Nov. 2004.
- R. Stark, J. H. Israel, and T. Wohler. Towards hybrid modeling environments—merging desktop-cad and virtual reality-technologies. *Annals of CIRP*, 59:179–182, Nov. 2010.
- R. Sukthankar, R. Stockton, and M. Mullin. Automatic keystone correction for camera-assisted presentation interfaces. In *Proceedings of International Conference on Multimodal Interfaces*, pages 607–614, Oct. 2000.
- D. Teeni. *Designs that fit: an overview of fit conceptualization in HCI*, chapter Designs that fit: an overview of fit conceptualization in HCI, pages 168–180. Human Computer Interaction And Management Information Systems: Foundations. M.E. Sharpe, 2006.

- M. P. Terence. Image-guided surgery and therapy: current status and future directions. In *Proceedings of The International Society for Optical Engineering*, pages 1–12, May 2001.
- K. Toma. 3D measurement of a surface point using a high-speed projector-camera system for augmented reality games. In *Proceedings of IEEE/SICE International Symposium on System Integration (SII)*, pages 84–89, Dec. 2010.
- S. Tungjitkusolmun, S. T. Staelin, D. Haemmerich, J.-Z. Tsai, H. Cao, J. G. Webster, F. T. Jr. Lee, D. M. Mahvi, and V. R. Vorperian. Three-dimensional finite-element analyses for radiofrequency hepatic tumor ablation. *IEEE Transactions on Biomedical Engineering*, 49(1):3–9, Jan. 2002.
- E. Turcoco and A. Verri. *Introductory Techniques for 3D Computer Vision*. Prentice Hall, 1998.
- G. Unal, A. Yezzi, S. Soatto, and G. Slabaugh. A variational approach to problems in calibration of multiple cameras. *Pattern Analysis and Machine Intelligence*, 29(8):1322–1338, Aug. 2007.
- S. Vogt, F. Wacker, A. Khamene, D. Elgort, T. Sielhrst, H. Niemann, J. Duerk, J. Lewin, and F. Sauer. Augmented reality system for mr-guided interventions: phantom studies and first animal test. In *Proceedings of SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Display*, Aug. 2004.
- F. K. Wacker, S. Vogt, A. Khamene, J. A. Jesberger, S. G. Nour, D. R. Elgort, F. Sauer, J. L. Duerk, and J. S. Lewin. An augmented reality system for mr image-guided needle biopsy: initial results in a swine model. *Radiology*, 238(2): 497–504, Feb. 2006.
- N. Wang, L. L. Huang, and B. Zhang. A fast hybrid method for interactive liver

- segmentation. In *Proceedings of Pattern Recognition (CCPR)*, pages 1–5, Oct. 2010a.
- Q. Wang, L. Fu, and Z. Z. Liu. Review on camera calibration. In *Control and Decision Conference*, pages 3354–3358, May 2010b.
- W. Wang, J. Hong, Y. P. Tang, and B. C. Shi. A multi-camera calibration technique based on active vision. In *Information Engineering and Computer Science*, pages 1–4, Dec. 2009.
- Z. Wang, B.-P. Nguyen, C.-K. Chui, J. Qin, C.-H. Ang, and S.-H. Ong. An efficient clustering method for fast rendering of time-varying volumetric medical data. *The Visual Computer*, 26(6-8):1061–1070, 2010c.
- P. Wen. Medical image registration based-on points, contour and curves. In *Proceedings of International Conference on Biomedical Engineering and Informatics*, pages 132–136, May 2008.
- R. Wen, C.-K. Chui, and K.-B. Lim. *Intraoperative visual guidance and control interface for augmented reality robotic surgery*, pages 191–208. *Augmented Reality-Some Emerging Application Areas*. InTech, 2010a.
- R. Wen, C.-K. Chui, and K.-B. Lim. Intraoperative visual guidance and control interface for augmented reality robotic surgery. In *2010 8th IEEE International Conference on Control and Automation*, pages 947–952, Jun. 2010b.
- J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *Pattern Analysis and Machine Intelligence*, 14(10):965–980, Oct. 1992.
- S. Wladyslaw and N. Artur. Filter-less gray patterns detection in 3d modeling by structured light. In *Proceedings of the Photonics Applications in Astronomy*,

- Communications, Industry, and High-Energy Physics Experiments (SPIE2009)*, pages 750203–1–750203–9, May 2009.
- X. D. Wu, X. H. Jiang, and J. X. Li. Review of traditional camera calibration methods in computer vision. *Journal of Fujian University of Technology*, 5(1): 57–61, May 2007.
- M. V. Wyawahare, P. M. Patil, and H. K. Abhyankar. Image registration techniques: An overview. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 2(3):1–27, Sep. 2009.
- T. Yamamoto, N. Abolhassani, S. Jung, A. M. Okamura, and T. N. Judkins. Augmented reality and haptic interfaces for robot-assisted surgery. *International Journal of Medical Robotics and Computer Assisted Surgery*, 8(1):45–56, Nov. 2005.
- J. Yamato, J. Ohya, and K. Ishii. Recognizing human action in time-sequential images using hidden markov model. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition(CVPR92)*, pages 379–385, Jun. 1992.
- L. Yang, R. Wen, J. Qin, C.-K. Chui, K.-B. Lim, and S.K.-Y. Chang. A robotic system for overlapping radiofrequency ablation in large tumor treatment. *IEEE/ASME Transactions on Mechatronics*, 15(6):887–897, Dec. 2010.
- Y. Yasumuro, M. Imura, Y. Manabe, O. Oshiro, and K. Chihara. Projection-based augmented reality with automated shape scanning. In *Proceedings of the Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments (SPIE2005)*, pages 555–562, Jun. 2005.
- M. L. Yuan, S.K. Ong, and A. Y. C. Nee. Technical section: A generalized registration method for augmented reality systems. *Computers and Graphics*, 29(6):980–997, Dec. 2005.

- S. Zhang and P. S. Huang. Novel method for structured light system calibration. *Optical Engineering*, 45(8):083601–083608, Apr. 2006.
- Z. Y. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, Nov. 2000.
- Q. Zhao. A survey on virtual reality. *Science in China Press*, 52(3):348–400, Mar. 2009.

List of Publications

The contents of this dissertation are based on the following manuscripts that have been submitted, accepted, or published by journals and conferences.

Journal Papers:

- [1] L. Yang, R. Wen, J. Qin, C.K. Chui, K.B. Lim and S.K.Y. Chang. A robotic system for overlapping radiofrequency ablation in large tumor treatment. *IEEE/ASME Transactions on Mechatronics*, 15(6), pages 887-897, Dec. 2010.
- [2] R.Wen, C.K. Chui, S.H. Ong, K.B. Lim and S.K.Y. Chang. Projection-based visual guidance for robot-aided RF needle insertion. *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, Accepted.

Book Chapter:

- [1] R.Wen, C.-K.Chui and K.-B. Lim, "Intraoperative visual guidance and control interface for augmented reality robotic surgery," in Augmented Reality-Some Emerging Application Areas, *InTech*, pp.191-208, 2010.

Conference Proceedings:

- [1] R. Wen, C.B. Chng, C.K. Chui, K.B. Lim, S.H. Ong and S. K. Chang, "Robot-assisted RF Ablation with Interactive Planning and Mixed Reality Guidance,"

IEEE/SICE International Symposium on System Integration (SII), Fukuoka, Japan, pages 31-36, Dec. 2012.

[2] R. Wen, L. Yang, C.K. Chui, K.B. Lim and S. Chang. Intraoperative visual guidance and control interface for augmented reality robotic surgery. in *Proc. IEEE Int. Conf. Control and Automation (ICCA)*, Xiamen, China, pages 947 - 952, Aug. 2010.

[3] C.K. Chui, C.B. Chng, T. Yang, R. Wen, W. Huang, J. Liu, Y.Su and S. Chang. Learning laparoscopic surgery by imitation using robot trainer. *IEEE International Conference on Robotics and Biomimetics*, Phuket Thailand, pages 947 - 952, Dec., 2011.