Research on Medical Image Processing Technologies for Fetal Surgical Navigation

胎児の手術ナビゲーションのための医用画像処理技術に関する研究

April, 2014

Graduate School of Global Information and Telecommunication Studies

Waseda University

Research on Image Processing II

Rong Xu

Contents

List of Figures

List of Table

Summary

Congenital diaphragmatic hernia (CDH), which has a prevalence of 1 in 2,000-3,000 newborns and leads to approximately 8% of the known major congenital anomalies, is a severe birth defect of the diaphragm. In recent clinical practice, it has been reported that severe CDH could be treated in utero by a minimally invasive surgery (MIS) called fetoscopic tracheal occlusion (FETO), which places a detachable balloon into the fetal trachea to prevent pulmonary hypoplasia by increasing the pressure of the chest cavity. In order to perform an FETO surgery, the surgeon inserts a fiber endoscope having a diameter of 1.3 mm within a cannula (Karl Storz) having a diameter of 3.3 mm into the amniotic cavity through the abdominal and uterine walls, towards the fetal mouth and trachea, being navigated by 2D ultrasound (US) images and fetoscopic images. However, this operation is so difficult and risky that high-level surgical skills are required.

To facilitate the FETO surgery, a flexible wire-driven surgical tool, whose posture can be transformed to fit the internal structure between the fetal oral cavity and trachea, has recently been developed by Zhang et al., but the operation of this surgical tool is still difficult without an automatic navigation system. A prospective FETO surgical navigation system should be able not only to guide the surgical tool to the fetal trachea by analyzing medical images such as 3D US images and fetoscopic images, but also to provide the surgeon and control module with feedback information obtained by the medical image analyses. Among different technologies needed for achieving the prospective surgical navigation system, this thesis focuses on the following three main topics:

(1) 3D fetal face detection from 3D US images;

(2) Surgical tool location from 3D US images and fetal mouth tracking in real-time from fetoscopic images;

(3) Estimation of fetal trachea's position in 3D tracking space (the world coordinate system of 3D electromagnetic (EM) tracking system).

As described below, new methods for the three topics are proposed and explored. In addition, basic technologies (a) to (c), which are needed for the above-mentioned (1) to (3), are dealt with by this thesis as follows:

(a) Medical image segmentation, which is a pre-process for detecting the fetal face and head by (1), and is a pre-process for (b) also;

(b) Ellipse detection, which is needed for detecting the fetal head by (1) and fetal mouth by (2), and is needed for (c) also;

(c) 3D US calibration, which obtains transformation matrix between 3D US image coordinate system and the EM world coordinate system, and is needed for (1) to (3).

The contents of each chapter of this thesis are as follows.

Chapter 1 describes the background, related work, purpose, outline of the proposed approach and organization of this thesis. The background overviews image-guided procedures (IGP), congenital diaphragmatic hernia (CDH) and its treatments including fetoscopic tracheal occlusion (FETO). The related work reviews IGP systems, medical imaging and image process, and current navigation systems for the FETO surgery.

Chapter 2 overviews some existing IGP systems, which include tomographic image overlay systems, fluoroscopic X-ray systems, CT and MR based systems, and video based augmented reality systems. Related technologies on medical image processing are illustrated, where technologies on enhancement, segmentation, feature extraction, registration, visualization, and human computer interaction are included.

Chapter 3 elaborates on the prospective surgical navigation system and its related technologies, where one section explains the technologies needed for the navigation system, and the other section explains about three main topics (1) to (3).

Chapter 4 proposes new methods for (a) to (c) and studies their validities experimentally, as follows: (a) An improved Kernel-based fuzzy C-means algorithm is proposed for more accurate image segmentation than conventional methods. Experiments using synthetic gray-scale images, simulated brain MR images, and fetal 2D US images show that the proposed method has a strong robustness against image noise, and also achieves a better segmentation performance than the conventional methods. (b) An improved iterative randomized Hough transform algorithm is proposed for accurate ellipse detection despite occlusions and discontinuities. Experiments using

clinical fetal US images demonstrate that the proposed method achieves more robust and accurate results, and has a better efficiency than conventional methods. (c) A novel 3D US calibration system using 24 fiducial points obtained by resin cones is proposed for more accurate calibration. Experimental results are comparable to those obtained by conventional methods using electromagnetic tracking.

Chapter 5 proposes a new method for (1). Specifically, a novel boosting traversal scheme based on the spatial relations between each facial feature in 3D US images is presented for detecting the fetal face and locating five key facial features, which involve nose tip, two eyes' inner corners, nose upper bridge and upper lip. Experimental results indicate the proposed approach is efficient and robust for 3D fetal face detection and feature location. The accuracy of locating the key features is 100% as a result of testing 72 images of 6 subjects, and the location error 3.18±0.91 mm of the detected upper lip can be tolerated by the FETO surgery. It turns out that the system could complete the entire detection process within approximately 625 msec on a quad-core 2.60 GHz computer.

Chapter 6 proposes new methods for (2): locating the surgical tool and tracking the fetal mouth. Each of the 15 points on the surgical tool is located in 3D tracing space (the world coordinate system of 3D EM tracking system) based on a 3D skeleton extraction technology and the 3D US calibration system. Experimental results show that an average TRE (target registration error) of 1.99±0.49 mm and a maximum TRE of 2.61±0.49 mm are achieved, which satisfy 3.85 mm, the requirement of the FETO surgery. The fetal mouth is tracked from a fetoscopic video sequence via the AdaBoost algorithm, the Camshift algorithm, and the improved iterative randomized Hough transform algorithm. Experimental results demonstrate that the proposed approach can automatically and accurately detect and track the fetal face and mouth in real-time in the fetoscopic video sequence.

Chapter 7 proposes a method for (3): estimating the position of the fetal trachea. First, a 3D fetal model with oral cavity and airways is reconstructed. Second, the 3D fetal facial surface is extracted from 3D US images. Third, the reconstructed 3D fetal model is registered with the extracted 3D fetal facial surface using sparse and dense matching, and is localized in the 3D tracking space based on the 3D US calibration system. Finally, the position of the trachea is estimated in the 3D tracking space based

on the geometric relationship between the internal organs and the 3D fetal facial surface in the 3D fetal model that can be known based on their design. In experiments, a mean TRE of 1.55±0.46 mm is achieved for the location accuracy of the 3D fetal facial surface extracted from 3D US images, and mean location errors of 2.51±0.47 mm and 3.04±0.59 mm are respectively achieved for the location accuracies of the pharynx and the trachea entrance, which satisfy the requirement of the FETO surgery.

Chapter 8 concludes this thesis and states future work.

Chapter 1 Introduction

1.1 Background

Recently, minimally invasive techniques are rapidly developed to replace traditional open surgical procedures in clinical practice because of their better results and lower overall costs. It is noteworthy that indirect, image-based feedback becomes popular instead of direct visual feedback, which is still challenging. In open surgery, the anatomical structures can be directly seen and felt by the physician. However, in image-guided procedures (IGP), the anatomical structures needs to be identified in the images by segmentation, and the spatial relationship between the images and the patient needs to be established by registration. Moreover, navigation accuracy of the image-guided procedures should be comparable or better than that accomplished by the traditional method. Therefore, the researches on recreating the spatial information and transferring a plan into action for image-guided procedures are more and more attracted.

For a better understanding of the spatial structure of the anatomy, image-guided systems frequently integrate medical images and other information, such as tracked instruments. These systems can contribute to the following aspects: (1) robustness of the final result for minimally invasive procedures can be improved, and the gap between exceptional and standard practice can be narrowed; (2) new minimally invasive procedures may be realized to allow physicians to carry out procedures that were formerly considered very dangerous; and (3) a quantitative comparison between plan and implementation can be allowed.

In medical procedures, physicians mentally integrate their knowledge of anatomical structures with specific medical images of the patient to produce a plan and perform it. Image-guided systems use a similar approach, where all information are incorporated and operated to provide guidance to the physician.

Initially, image-guidance systems were accepted by two medical procudures, neurosurgery [1-3] and orthopedics [4, 5], because both can accept a rigid anatomy assumption. Furthermore, the rigid anatomy assumption is also applied for all current commercial image-guided systems. Deformable anatomical structures are still studied for image-guided systems [6].

1.1.1 Image-guided Procedures

Image-guided procedures can be divided into three phases based on a time-line view: pre-operative, intra-operative, and post-operative. Each phase will be described as follows.

- **Pre-operative**

The goal of this phase is to make a surgical plan according to pre-operative images and additional information. Multiple imaging modalities are often used concurrently, such as CT (computed tomography) and MRI (magnetic resonance imaging) for optimally visualizing bone and soft tissue. All data are mapped in one coordinate system during plans formulation. These plans are relative to specific patient and procedure, and thus have variable complexity [7].

The key technologies on medical image processing involved in pre-operative phase are: (1) enhancement, the improvement of the image quality, (2) segmentation, the classification of meaningful anatomies in the images; (3) registration, the transformation of different sets of data into a single coordinate system; and (4) visualization, the display and manipulation of image and data;

- **Intra-operative**

During the operation, the coordinate system specified for the plan must be mapped to a coordinate system in the operating room (OR). The image-guided system can visualize tracked tools and anatomies and display their spatial relationships for the physician, using all sorts of interfaces and displays. Additionally, the anatomical picture may be updated by intra-operative images. If necessary, the operation plan is then revised according to the surgical progress. The location of the tools and the anatomy can be recorded by the guidance system, and can be used for intervention assessment.

The key technologies involved in intra-operative phase cover all the technologies involved in pre-operative phase, in addition to: (1) tracking systems, for locating the anatomies and tools; and (2) human computer interaction (HCI), for interacting with the guidance system and visualizing the information to the physician.

- **Post-operative**

In post-operative phase, the acquired images and the results are used to make a

quantitative comparison with the pre-operative plan.

The key technologies used in post-operative period are the same as those used in pre-operative period.

Motivated by similar reasons of IGP systems and the requirement of a higher security and operability, this thesis deals with medical image processing technologies that could contribute to achieving an automatic surgical navigation system for guiding fetoscopic tracheal Occlusion (FETO) surgeries, which try to treat a severe birth defect called congenital diaphragmatic hernia (CDH). So far there has been no automatic IGP system for such a minimally invasive surgery.

1.1.2 Congenital Diaphragmatic Hernia (CDH)

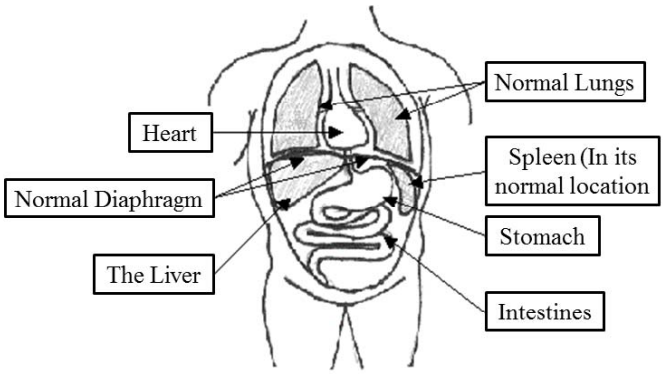
Congenital diaphragmatic hernia (CDH) is a birth defect of the diaphragm, which occurs sporadically, with an incidence of 1/2000 to 1/5000 of newborns, depending on whether stillbirths are included. CDH allows the abdominal organs to move into the chest cavity and cause two complications: pulmonary hypoplasia and pulmonary hypertension [8]. Survival of those fetuses with isolated CDH after postnatal surgery is about 50%. In recent clinical practice, it has reported that severe CDH could be treated appropriately in uterus by a minimally invasive surgery (MIS) called fetoscopic tracheal occlusion (FETO) [9], in which a detachable balloon is placed via the larynx into the fetal trachea to prevent pulmonary hypoplasia or pulmonary hypertension by increasing the pressure of the chest cavity.

(1) Introduction of CDH

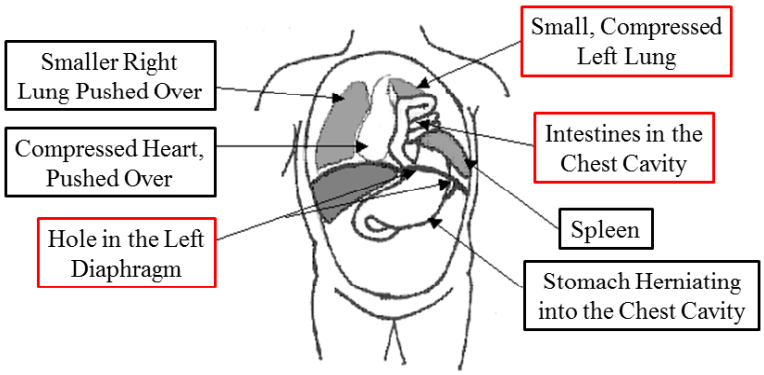
The diaphragm is a thin layer of muscle, and separates the chest and abdominal cavity. It is the muscle that the body uses to breathe, as shown in Fig. 1.1 (a). CDH means that there is a hole in that layer - or, rarely, that the diaphragm is missing altogether. A severe CDH can allow the organs, such as the stomach, intestines, and liver, to slip into the chest through the hole, which affects the normal growth of the lung on that side, as shown in Fig. 1.1 (b). Due to no full-grown lung, the fetus will have serious breathing troubles after he/she is born.

Eighty-four percent of the hernias occur on the left side, 13% on the right and only 2% are bilateral. CDH will cause problems on the growth of both airway and vessels,

and thus newborns will experience severe respiratory insufficiency depending on the degree of pulmonary hypoplasia. Therefore, it is very critical to assess this anomaly and manage CDH in the perinatal period [10, 11].



(a) Normal Diaphragm



(b) A left-sided diaphragmatic hernia

Figure 1.1 Normal diaphragm and a left-sided diaphragmatic hernia. (Image provided courtesy of Texas Pediatric Surgical Associates, and they are linked by http://www.pedisurg.com/pteduc/ congenital_ diaphragmatic_ hernia.htm).

(2) History of Treatments of CDH

CDH was initially treated by anatomical repair [12]. Tracheal occlusion (TO) was applied to trigger lung growth by preventing egress of lung fluid and increasing airway pressure, which is proved by several research groups [13]. The Philadelphia group [14] reported a survival rate of 33% by TO. The San Francisco team [15] moved to the uterus by endoscopy, but laparotomy was still required to expose uterine, and there was a high risk of damage to internal organs. Endoscopy was first used by Deprest et al [16] to position balloon into the trachea under ultrasound (US) guidance [17], but a lower success rate was achieved [18, 19]. Subsequently, endoluminal plugging was first

reported in active fetal surgery programs in the United States either by laparotomy or percutaneously [20, 21]. In Europe, the clinical trials on severe cases were started by the Fetal Endoscopic (or Fetoscopic) Tracheal Occlusion (FETO) Task Force [22, 23].

1.1.3 Fetoscopic Tracheal Occlusion (FETO)

Today, the procedure of fetoscopic tracheal occlusion (FETO) is believed to work because it prevents egress of lung fluid, increasing airway pressure, causing proliferation, increased alveolar airspace and maturation of pulmonary vasculature [13]. During the operation, the balloon is inserted at 26–28 weeks and removed at 34 weeks, at least in severe cases [24].

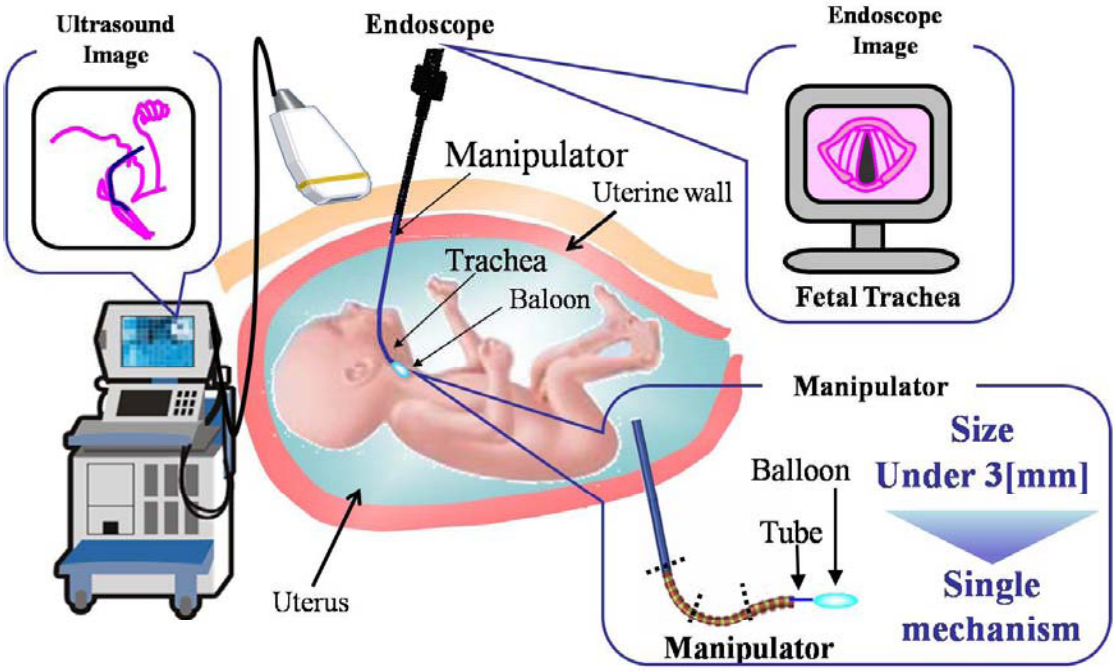


Figure 1.2 FETO surgery.

In an FETO surgery, the surgical tool with a fetoscope is inserted through the abdominal and uterine walls into the amniotic cavity, and directed towards the fetal mouth, and a detachable balloon is placed above the fetal trachea for the tracheal occlusion, as shown in Fig. 1.2, to stimulate lung growth by increasing pressure in the fetal chest cavity. During the FETO surgery, the fetal and placental position is determined by ultrasound examination for optimal trocar insertion. Sometimes, the fetal position is adjusted to get better access to the trachea by gentle external manipulation.

Once the desired position is achieved, some drugs will be injected to provide fetal anesthesia, immobilization and prevention of fetal bradycardia.

The endoscope is introduced into the fetal mouth and directed over the tongue in the midline by visualizing the raphe of the palate and uvula, to the pharynx, then the larynx with the epiglottis as a landmark and finally through the vocal cords to the trachea (Fig. 1.3). Then, the carina is identified by the endoscope, and the catheter is positioned to deliver the balloon just above it [22].

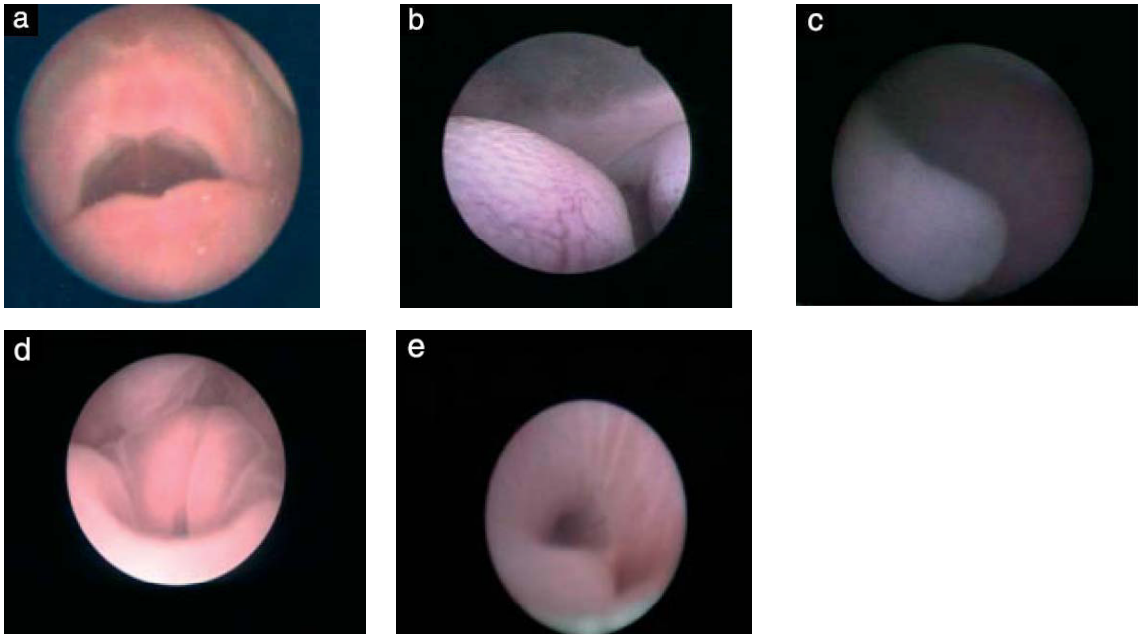


Figure 1.3 The endoscope is inserted into the fetal mouth and the raphe of the palate (a), the tongue (b), uvula and epiglottis (c) and vocal cords (d) are used as landmarks to gain access to the trachea. The endoscope is then advanced to identify the carina (e) and the balloon is placed just above it (Image provided courtesy of J. Deprest, et al.).

1.1.4 A Novel Flexible Robotic Manipulator (Surgical Tool)

Due to lack of flexibility, most surgical tools used in current FETO surgeries easily hurt fragile fetal internal organs during the operation, especially the position of the wound marked in Fig. 1.4. In order to reduce the possibility of such hurts, instead of the cannula used in FETO surgeries, a flexible wire-driven manipulator (the surgical tool to be used for FETO surgeries) with 6 degree of freedom (DOF) is designed and developed by Zhang, et al. [25, 26], where its outer diameter is 2.4 mm and inner diameter is 1.0 mm, as shown in Fig. 1.5.

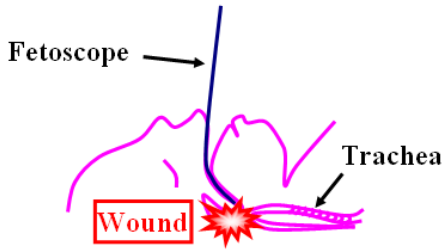


Figure 1.4 Risk of current methods.

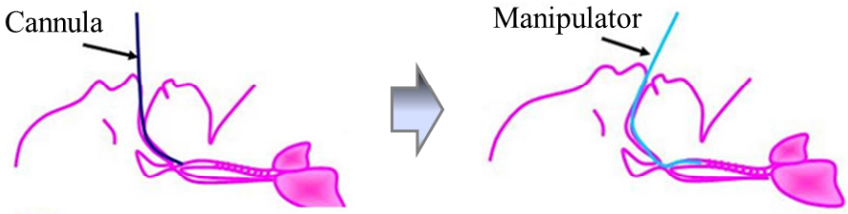
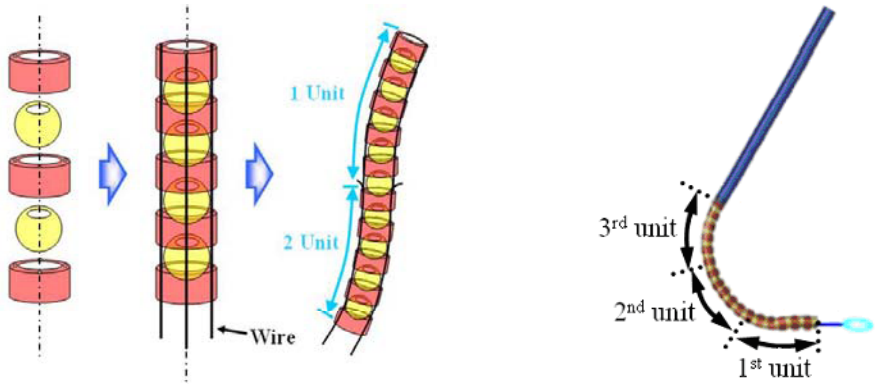


Figure 1.5 Improvement from the cannula to the designed manipulator.



(a) (b)

Figure 1.6 Multi-unit bending mechanism (a) and 3 units (b) of the manipulator.

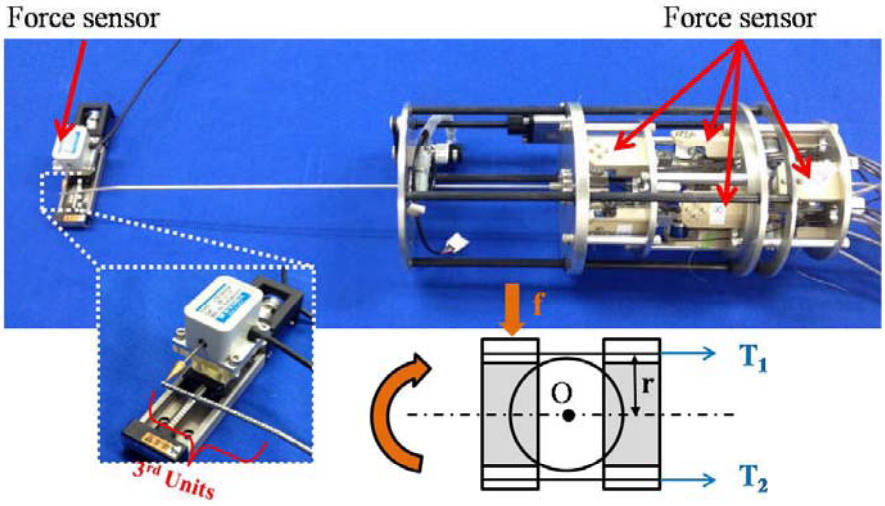


Figure 1.7 Robot control system of the manipulator.

Figure 1.6 (a) shows multi-unit bending mechanism of the manipulator. The final manipulator to be used for FETO surgeries consists of three units, and each containing 10 ball-joint-shaped arthroses, as shown in Fig. 1.6 (b). The robot control system of the manipulator is shown in Fig. 1.7, whose actions are operated by a computer.

1.1.5 Prospective Surgical Navigation System

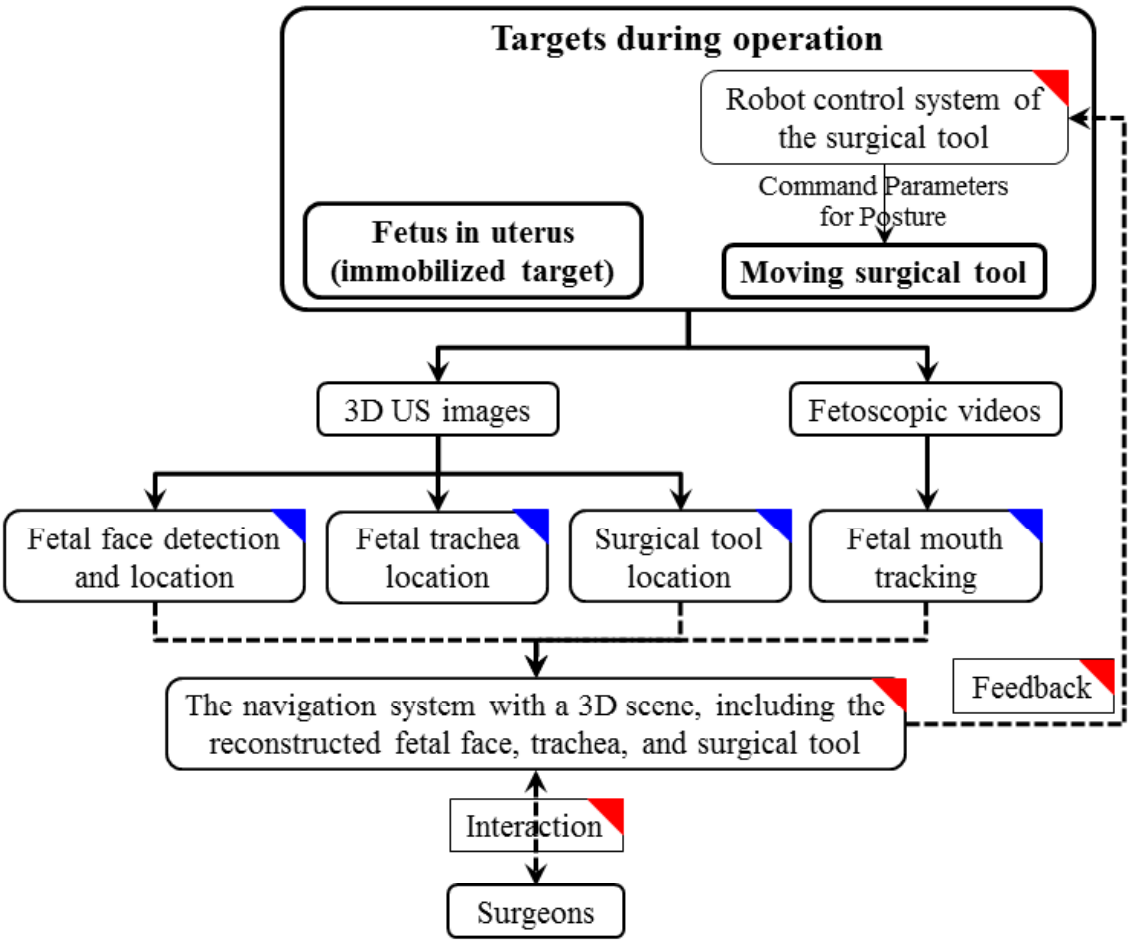


Figure 1.8 Prospective surgical navigation system for the FETO surgery. Blue triangle: topics studied by this thesis; red triangle and dotted line: processes and items not studied by this thesis.

During the current FETO surgery, surgeons should be able to reconstruct 3D structures of and to grasp the spatial relationship between the moving surgical tool, the fetus in uterus and his/her critical internal organs by watching the US images and fetoscopic images based on their clinical experiences. However, the reconstructed surgical scenes in the surgeons' brains are often not accurate and reliable. Therefore, a

prospective surgical navigation system that can pursue accuracy, efficiency, safety and reliability of the FETO surgery can be designed as shown in Fig. 1.8. The navigation system should be able to provide precise, accurate, timely, robust and reliable reconstructions that can be visualized as 3D scenes to be displayed. This facilitates the surgeon's vision based various judgments. In addition, the navigation system should be able to support difficult tasks for the surgeon. For example, the insertion of the flexible surgical tool into the fetal trachea is a very difficult and risky task for the surgeon, but if the navigation system can automatically or semi-automatically perform this task by detecting the fetal face, surgical tool, fetal mouth and estimating the position of the fetal trachea (Fig. 1.8), it is very helpful and useful for the surgeon. Of course, important judgments such as checking whether the surgical tool is at a proper position with the proper posture should be done by the surgeon based on the information and data provided by the navigation system. Namely, the system should promise rapid, timely and accurate responses and feedback to the surgeon during surgeries through reasonable human computer interaction.

This thesis deals with some image processing technologies that can contribute to supporting the surgeon's task for inserting the flexible surgical tool into the fetal trachea. The specific topics this thesis treats are indicated by blue triangles in Fig. 1.8. More details about each topic are described in Section 1.4 .

1.2 Related Work

1.2.1 IGP Systems

There are some existing IGP systems to be assisted by medical image processing technologies, where (1) Tomographic image overlay systems used technologies on medical image processing and visualization; (2) Fluoroscopic X-ray systems used technologies on medical image processing, visualization, and tracking systems; (3) CT/MR based systems used technologies on medical image processing, visualization, segmentation, registration, and tracking systems; (4) Video based augmented reality systems used technologies on medical image processing, visualization, segmentation, registration, and tracking systems. Each system is detailed in Chapter 2 .

1.2.2 Medical Imaging and Image Processing

Medical images are the main information source of image-guided procedures. In the first medical procedure performed in 1896, X-ray images were used [27, 28]. The common medical imaging modalities are computed tomography (CT), magnetic resonance imaging (MRI), X-ray, functional MRI (fMRI), positron emission tomography (PET), single photon emission computed tomography (SPECT), X-ray fluorocopy, C-arm CT, ultrasound (US), optical imaging, etc. More details about each modality can be found in [6].

On the other hand, as the key technologies of IGP systems, techniques of medical image processing can be divided into five categories: enhancement, segmentation, registration, visualization, and human computer interaction (HCI). Each category includes a variety of solutions for different medical imaging modalities and distinct clinical applications.

(1) Enhancement

Enhancement algorithms are used to reduce image noise and increase the contrast of structures of interest. In medical images, the boundaries between each tissue are not obvious, and thus accurate interpretation may become difficult if the images contain much noise. Usually, enhancement can improve the image quality and facilitate diagnosis. A clearer image can generally be provided for a human observer by enhancement techniques. These techniques can also be utilized as a pre-processing step for subsequent automated analysis.

(2) Segmentation

Segmentation is the stage to identify structures of interest from the images, which is still challenging in algorithm implementation. In most cases, the final result of the entire image analysis depends on the segmentation approach, because measurements and other processing steps are based on segmented regions. There are some typical segmentation algorithms, such as thresholding, region growing, deformable templates, and pattern recognition techniques such as neural networks and fuzzy clustering, where fuzzy clustering methods have been more widely used for medical image segmentation.

(3) Registration

Registration of two images of the same part of human body is essential for many applications, where the desired information can be conveyed by the correspondence between the two images. Using different modalities often require registration process, for example CT and MRI, can be taken from the same patient with the same instrument at different times, or can belong to two different subjects. Anatomic atlas templates or models also require registration algorithms for comparison with acquired images. These algorithms must be able to deal with the distortions between the two images, which may be caused by differences between the imaging modalities, artifacts, soft tissue elasticity, and variability among subjects.

(4) Visualization

Visualization is a significant tool for medicine and biology. For the visual systems, graphics hardware and software are specifically designed to facilitate visual inspection of medical and biological data. For volumetric data, visualization techniques are essential to enable effective visual inspection [29]. In this thesis, surface rendering [30, 31] that is the most popular rendering method in computer graphics is applied in this thesis to visualize 3D objects.

(5) Human Computer Interaction

For IGP systems, computers are commonly used in the operating room (OR), and thus human computer interaction (HCI) is also required during operation. Interaction requirements are dynamic in the OR based on type of the operation, surgeon's decision, and patient profile, etc. An HCI designing framework should consider the main intro-operative aspects, such as surgical workflow, OR human roles, and interaction requirements with hardware and software solutions. According to HCI, surgeons can communicate with IGP systems by some input components, such as mouse/keyboard, Kinect, and inertial sensors, etc.

More reviews of each category of the technologies are stated in Chapter 2 .

1.2.3 Current Navigation Systems and Issues to Be Solved

In conventional navigation systems for the FETO surgery, real-time 2D or 3D US

images are applied simply to visualize the position of the surgical tool and fetal mouth, without any further processing of US images or an accurate location of the target organs [22, 32, 33]. Moreover, since the surgical tool and fetal mouth are still difficult to clearly distinguish using 2D or 3D US images because of speckle noise, shadows, and a low signal-to-noise ratio, surgeons are required to have a wealth of experience in using US images. The group of Deprest examined US images to determine the fetal and placental position for optimal trocar insertion, and guide a cannula with a purpose-designed trocar to insert into the amniotic cavity to the fetal mouth. In succession, the fetoscope was introduced into the fetal mouth as a guide and advanced over the tongue to the pharynx, then to the larynx with the epiglottis, and finally through the vocal cord to the trachea [22]. Ruano et al. reported a guidance of 4DUS to direct a cannula with the fetoscope into the fetal oral cavity, which allowed imaging of the entire fetal face and the insertion of the cannula toward the fetal oral cavity [32]. Tchirikov [33] also discussed real-time 3D US guidance in FETO surgery to reduce the surgical risk.

However, none of these works achieved automatic and accurate IGP systems for FETO surgeries; in other words, none can actualize the prospective surgical navigation system shown in Fig. 1.8. To automatically navigate the flexible surgical tool developed by Zhang, et al. into the fetal trachea, the following issues need to be solved.

(1) 3D fetal face detection from 3D US images;

(2) Surgical tool location from 3D US images;

(3) Fetal mouth tracking from fetoscopic images;

(4) Estimation of the fetal trachea's position from 3D US images;

(5) Feedback to the robot control system of the surgical tool;

(6) Interaction between the surgical navigation system and surgeons.

Among the above-mentioned issues, this thesis focuses on the issues (1) to (4), all of which could be solved by medical image processing technologies. More specifically, this thesis aims at solving the issues on the fetal face detection from 3D US images, surgical tool location from 3D US images and fetal mouth tracking from fetoscopic images, and estimation of the fetal trachea's position from 3D US images.

1.3 Purpose of the Thesis

The purpose of this thesis is to solve the four issues described in Section 1.2.3. Specifically, this thesis proposes and explores the four methods indicated by the blue triangle in Fig. 1.8. The goal of each of the four methods is as follows, where (2) and (3) in Section 1.2.3 are combined into one item, and thus three main topics are as follows.

(1) 3D fetal face detection

Robust, efficient detection of fetal faces should be achieved from 3D US images with high accuracy: i.e. face location errors should be within the tolerance determined by the FETO surgery's requirement.

(2) Surgical tool location and fetal mouth tracking

Accurate location of the flexible surgical tool should be achieved from 3D US images, where the location accuracy should be within the tolerance determined by the FETO surgery's requirement.

Accurate, real-time tracking of the fetal mouth should be achieved from fetoscopic video sequence.

(3) Estimation of the fetal trachea's position

The position of the fetal trachea should be estimated from 3D US images, where the accuracy of the position estimation should be within the tolerance determined by the FETO surgery's requirement.

To achieve the above-mentioned three goals, this thesis needs to propose the following three basic technologies: (a) medical image segmentation, (b) ellipse detection, and (c) calibration for 3D US images. The relationship between the above-mentioned three main topics (1) to (3) and the three basic technologies (a) to (c) is indicated in Fig. 1.9, where (a) is used by (1) and (b), (b) is used by (1), (2) and (c), and (c) is used by (1), (2) and (3). The goals of (a) to (c) are as follows.

(a) Medical image segmentation

MR and US images should be segmented accurately. In particular, the fetal head and the flexible surgical tool should be segmented from the background in US images.

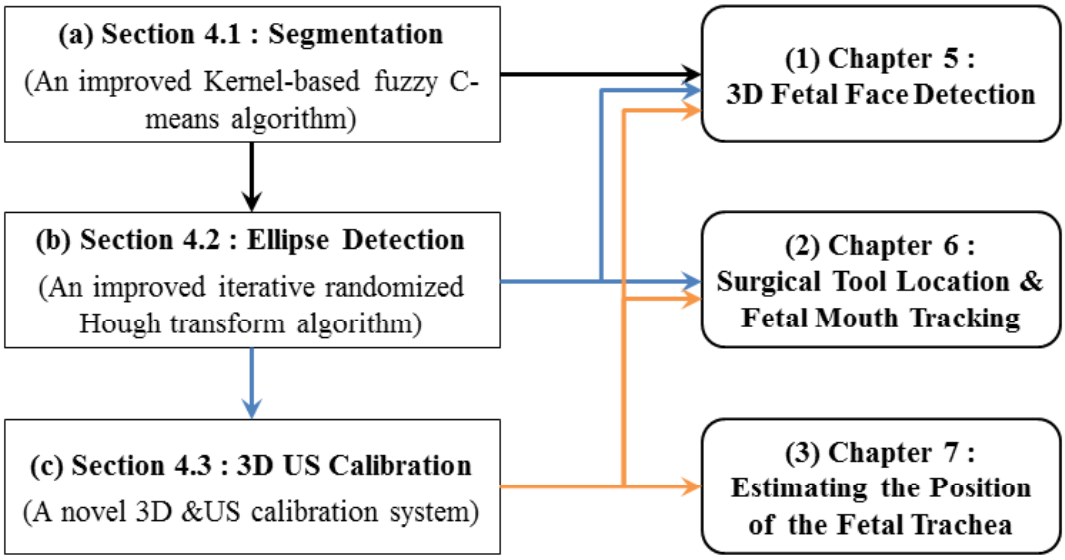


Figure 1.9 Relationship between the basic technologies and the three main topics.

(b) Ellipse detection

Partially occluded or discontinued ellipses, which could correspond to fetal heads in US images and fetal mouth in fetoscopic images, should be detected accurately. As explained in later sections, cones are used for (c) 3D US calibration; therefore, ellipses, which correspond to the cones, should be detected accurately.

(c) 3D Ultrasound Calibration

The coordinate system of the 3D US images, the world coordinate system of the 3D tracing space, and local coordinate systems of the fetuses and surgical tool should be calibrated in an accurate, robust manner.

1.4 Proposed Approach

For the FETO surgery, (1) 3D fetal face detection, (2) surgical tool location and fetal mouth tracking, and (3) estimation of the fetal trachea's position, play an important role on reducing the surgical risk, as shown in Fig. 1.2. Therefore, the prospective surgical navigation system for the FETO surgery is planned in Fig. 1.8, where during operation, real-time 3D US images are used for locating the fetal face, trachea, and surgical tool in a single coordinate system. Simultaneously, fetal mouth is tracked in real-time from fetoscopic images, where the fetospcopy is supposed to be inserted into

the surgical tool and moves together with it. Finally, all of the above obtained information can be applied as a feedback to the robot control system of the surgical tool for guiding its movement. This thesis proposes a novel approach for each of the three topics (1) to (3) so that the corresponding three goals (1) to (3) described in Section 1.3 are achieved.

(1) 3D Fetal Face Detection

3D Fetal face detection approach is applied to detect fetal face and facial key features from 3D US images, where (1-1) the detected fetal face and key features are located in 3D tracking space (the world coordinate system of 3D electromagnetic (EM) tracking system) based on the 3D US calibration system, and then their locations can be used as a feedback to the robot control system; (1-2) the detected key features are also used for the proposed ICP-based registration method (Section 7.3).

(2) Surgical Tool Location and Fetal Mouth Tracking

To locate the surgical tool, first 3D skeletons of the surgical tool are extracted at first from 3D US images, and then located in 3D tracking space based on the 3D US calibration system. Moreover, fetal mouth tracking from fetoscopic videos is realized in real-time based on the ellipse detection method described in (b). Both can be used as a feedback to the robot control system in the future.

(3) Fetal Trachea Location

Although 3D US images can allow 3D modeling of the surface of a fetus, they cannot 3D-model internal organs of a fetus, because the internal organs cannot be identified in 3D US images. Meanwhile, MR images are available to 3D-model internal organs of a fetus in actual clinical applications, but in case of this thesis, it is very difficult to get a permission to use real fetal MR images due to privacy problems in Japan. Therefore, as an alternative for MR images, this thesis uses models of fetal phantoms with internal structure. The model is registered to the 3D US images so that the internal organs including trachea are located in the 3D US images. Based on the 3D US calibration system described in (c), fetal trachea is also located in the 3D tracking space, which can be used as a feedback to the robot control system in the future.

Note that in the prospective surgical navigation system in Fig. 1.8, the modules

(topics) with the blue triangle correspond to the above-mentioned (1) to (3). In contrast, the modules with the red triangle are included in the future work, where examples of such modules involve displaying 3D scenes with the reconstructed fetal facial surface, trachea and surgical tool, human computer interaction for surgeons, and the feedback to the robot control system of the surgical tool.

As described in Section 1.3 , this thesis aims at achieving the goals of the three basic technologies (a) to (c). The approaches of the technologies (a) to (c) are as follows.

(a) Medical image segmentation

An improved Kernel-based fuzzy C-means algorithm is proposed for segmenting medical images such as MR and US images.

(b) Ellipse detection

An improved iterative randomized Hough transform (IRHT) algorithm is proposed for detecting partially occluded or discontinued ellipses from US images and fetoscopic images.

(c) 3D Ultrasound calibration

A novel 3D ultrasound calibration system based the 24 fiducial points obtained by the resin cones placed in the area scanned by a 3D US probe is proposed so that the transformations between different coordinate systems can be estimated.

1.5 Organization of the Thesis

As shown in Fig. 1.10, this thesis is organized as follows.

Chapter 1 is the introduction of this thesis.

Chapter 2 elaborates on conventional related works such as some existing IGP systems and video based augmented reality systems, as well as medical image processing technologies such as enhancement, segmentation, feature extraction, registration, visualization, and human computer interaction.

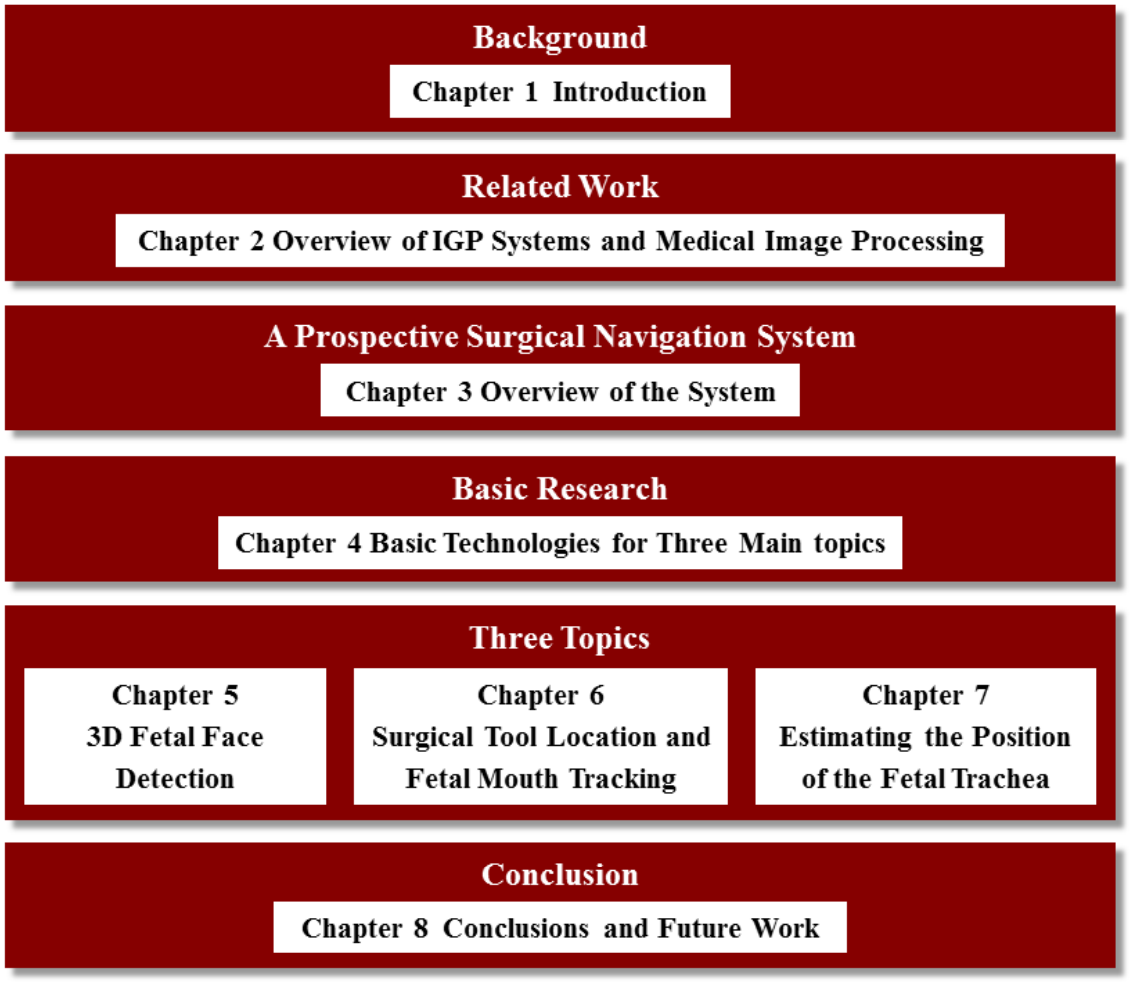


Figure 1.10 Organization of the thesis.

Chapter 3 explains the prospective surgical navigation system and overviews the three proposed methods: (1) 3D fetal face detection, (2) surgical tool location and fetal mouth tracking, and (3) fetal trachea location, where (1) to (3) are explained in Section 1.4 .

Chapter 4 proposes and explores the three basic technologies: (a) medical image segmentation, (b) ellipse detection, and (c) 3D Ultrasound calibration, which are described in Section 1.4 . The approaches of the three basic technologies are explained. Experimental results demonstrate the validity of the proposed approaches.

Chapter 5 proposes a method for detecting 3D fetal faces. The specific algorithm of the proposed method is explained. Experimental results show that this process satisfies the goal.

Chapter 6 proposes methods for surgical tool location and fetal mouth tracking. Their specific approaches are explained. Experimental results show the effectiveness of

the proposed approaches.

Chapter 7 proposes a method for fetal trachea location. The specific approach is explained. Experimental results demonstrate the validity of the proposed method.

Chapter 8 concludes this thesis and states future work.

Chapter 2 Overview of IGP Systems and Medical Image Processing

2.1 Some Existing IGP Systems

As introduced in Section 1.2.1, details of those existing IGP systems are described in this section.

2.1.1 Tomographic Image Overlay Systems

A new image display approach with semi-transparent mirrors is developed for displaying US images in real-time [34], CT [35] and MR [36] images, as shown in Fig. 2.1, but these tomographic image overlay systems have not been commercially applied. The physician views the patient through a mirror or a semi-transparent mirror, and the image reflection from the monitor appears to float inside the patient. These systems are the most suitable for needle based procedures because one tomographic slice is sufficient for navigation.

2.1.2 Fluoroscopic X-ray Systems

For fluoroscopic X-ray systems (also called virtual fluoroscopy), a fixed set of X-ray images and dynamic, real-time, projections of tool representations onto the images are utilized to replace C-arm fluoroscopic X-ray imaging [37, 38].

Figure 2.2 shows an example of the Medtronic system. These systems are primarily used for qualitative planning, or when immediate intervention is required (i.e. trauma), because no pre-operative imaging is obtained. Additionally, all objects are assumed to be rigid and are either dynamically tracked by optical tracking or immobilized. These systems were gladly approved by the orthopedics community because their workflow only diverges slightly from current practice.

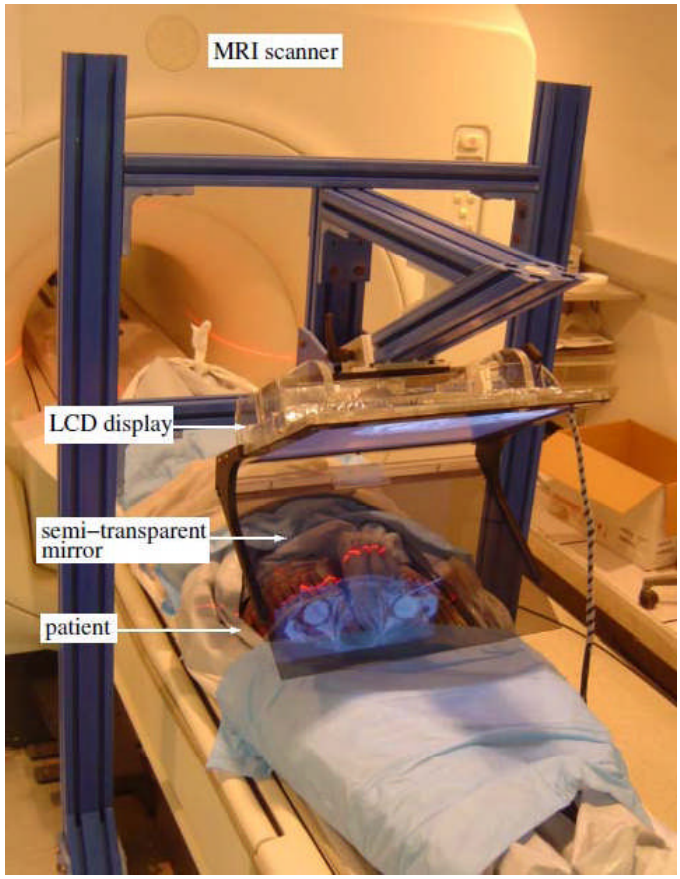


Figure 2.1 MR based tomographic image overlay guidance system (Image provided courtesy of G. Fichtinger).

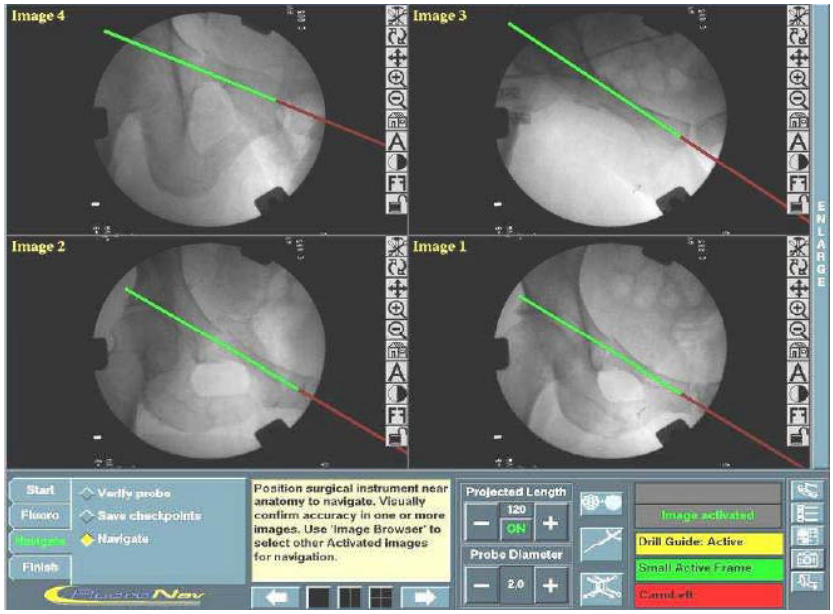


Figure 2.2 Screenshot of the Medtronic Fluoronav® virtual fluoroscopy system. A tool and its virtual extension are projected onto the image set. (Image provided courtesy of Ziv Yaniv, et al.).

2.1.3 CT and MR based Systems

CT and MR-based systems use dynamic, real-time, 3D views of anatomical structures and surgical tools to replace stereotactic frames. All tools are required to be fitted with reference frames so that they can be tracked in real-time, and anatomies are either fitted with a dynamic reference frame and tracked, or immobilized. The information is displayed via the quadrant view for most systems [39]. These systems have been commercially applied for procedures of rigid anatomical structures, while deformations can also be dealt with by some research systems, such as brain shift [40].

2.1.4 Video based Augmented Reality Systems

For video-based augmented reality systems, the video with information about the anatomical structures is overlaid with the tools that are underneath the visible surface. These systems combine all the technical aspects of virtual fluoroscopy systems and MR/CT based systems. Specifically, navigation that is based on 2D projective images is similar to virtual fluoroscopy, but registration is similar to CT/MR based systems. Information is displayed either on a screen or directly into an optical device (e.g. operating microscope) for viewing the anatomy. Conveying depth information is still an issue, because navigation is based on 2D projective images [6].

2.2 Medical Image Processing

2.2.1 Enhancement

Medical images are often accompanied by noise due to various sources of interference and different imaging and data acquisition systems, which reduce the contrast and the visibility of details. Therefore, it may be helpful for a medical specialist to interpret an image if its appearance and visual quality are improved, even if only subjective.

A given image can be improved by image enhancement techniques so that it can become easier to comprehend desired image features by the human visual system or more likely to be detected by automated image analysis systems [41, 42]. Basic

techniques of image enhancement include histogram equalization, mean and median filtering, edge enhancement, image averaging and subtraction, and the Butterworth filter [43]. Some advanced techniques for image enhancement involve the adaptive Wiener filter [44], nonlinear contrast enhancement techniques, a wavelet-based framework, a hybrid filter incorporating Fourier descriptors, etc. Each method serves a specific need and has its own realm of applications [29].

In this thesis, the used ultrasound images frequently contain much speckle noise due to its inherent characteristic, and thus operations of enhancement and filter on US images are particularly critical for subsequent processing. Herein, the used technologies of enhancement include Gaussian filter, white top-hat transform, morphological operations (opening and closing), bilateral filter, etc.

2.2.2 Segmentation

For IGP systems, segmentation is defined to separate the image domain into non-overlapping connected regions, which correspond to different anatomies. While medical image segmentation has been studied extensively [45, 46], there is still no automatic technique for imaging modalities and anatomies [47]. Therefore, the quality of segmentation results only can be judged by the physician.

The most common algorithms include thresholding, watershed methods, edge-based methods, gradient operators-based methods, region-based methods, probabilistic Bayesian method, Markov random field model, level set methods, atlas-based methods, statistical appearance models, deformable models, pattern recognition techniques such as neural networks and fuzzy clustering, and volumetric segmentation, etc. Several general surveys on medical image segmentation can be found in [45, 46], and specialized surveys on deformable models [48-50], vessel extraction [51-53], and brain segmentation [54-56] are also available.

Automated segmentation of medical images is a hard task, since images are often noisy and often contain multiple anatomical structures, and sometimes the organ boundaries may be confused. To overcome these challenges, domain-specific prior knowledge is integrated by many algorithms [57-59].

Among segmentation technologies, fuzzy c-means (FCM) algorithm [60] is a typical clustering algorithm, which uses Euclidean distance to measure similarity among

image pixels, but it is just effective in clustering 'spherical' data. In order to cluster more general dataset, a kernel-based fuzzy c-means (KFCM) algorithm [61, 62] is proposed by exploiting kernel function to measure data's similarities instead of Euclidean distance. However, KFCM method is easily affected by noise owing to ignoring spatial information in images. Recently, numerous methods [63-68] have been proposed by exploiting image spatial correlation to advance segmentation capability for low signal-to-noise ratio (SNR) images. In an image, the pixels should be highly relevant, i.e. the pixels in their immediate neighborhoods should possess almost the same characteristics. That is why, the spatial constraints among neighboring pixels is an important feature that can be contributing to image segmentation. In order to reduce the effect of noise, an improved KFCM method (iKFCM) [69] is proposed by cooperating with spatial information to optimize the objective function in conventional KFCM method in this thesis.

2.2.3 Feature Extraction

As a technology of feature extraction, ellipse detection is a specific solution for the prospective surgical navigation system, since position of fetal head can be detected an ellipse from 2D US images. Initially, approaches like least squares fitting [70, 71] and Hough transform (HT) [72] were the main approaches for this purpose. To overcome some limitations of the HT, combinatorial Hough transform (CHT) [73], randomized Hough transform (RHT) [74], probabilistic Hough transform (PHT) [75], dynamic generalized Hough transform (DGHT) [76], and Random Sample Consensus (RANSAC) [77] algorithm were proposed to improve the performance of HT for non-linear problems like ellipse detection. However, both RHTs and RANSAC may fail when strong noise can corrupt the curve-related peaks in the parameter space, or can generate a larger consensus set for a false model instance. A novel method named iterative randomized Hough transform (IRHT) [136, 137] was proposed for detection of partial ellipses under strong noise conditions. However, there are three limitations for IRHT algorithm: (1) all extracted edge pixels that cannot be grouped for one elliptic hypothesis are used for ellipse detection; (2) random selection of pixels for parameter estimation, which both limitations require a large number of samples and result in a low detection efficiency; (3) the robustness of the algorithm is not strong, because it does not consider the number of pixels on the ellipse. In order to overcome the limitations of

the original IRHT method, an improved iterative randomized Hough transform algorithm is proposed in this thesis.

On the other hand, in order to locate the fetal facial surface for the prospective surgical navigation system, key facial features, e.g. nose, eyes, lips and mouth, are necessary to be detected from 3D US images. Some earlier works on curvature analysis [78-81] are proposed for face recognition based on distinct facial curvature features. 3D model-based methods [82, 83] are often employed for face recognition by fitting a priori 3D face model to a given face. In addition, some methods are depicted to represent the 3D data in a different domain for face recognition, e.g. 3D PCA [84], shape index [85], point signature [86], and spine image [87], etc. Some other methods, e.g. face profile extraction [88], and facial curves [89], are also reported recently. However, face recognition technologies are rarely exploited in medicine, especially for 3D US images. To our best knowledge, there is only one work to detect 3D fetal face from US volumes for the optimal view [90], but their method required a priori 3D face mesh model for face detection and a high calculation complexity on the registration between the 3D model and the object. As a result, a curvature feature-based approach is proposed in this thesis to automatically detect 3D fetal face and accurately locate key facial features for guiding the FETO surgery in a high efficiency [180].

2.2.4 Registration

Image transformations that build correspondence between points or regions within images, or between physical space and images, can be estimated by registration algorithms. Thus, registration technologies can be divided into several classes: landmark/point-based registration [91], surface matching (e.g., the “Head and Hat” algorithm [92], distance transforms [93], iterative closest point [94]), registration based on voxel similarity measures [95] and minimizing intensity difference [96], correlation techniques [97], ratio image uniformity (RIU) [98], partitioned intensity uniformity (PIU) [99], information theoretic techniques (e.g., joint entropy [100], mutual information [101]), 2D-3D registration, and non-rigid registration, etc.

The first application of medical image registration was image-guided procedure. There are three factors that should be considered to evaluate the performance of a registration for image-guided procedures: (1) accuracy, measured by the target registration error (TRE) [91], which indicates how far the estimated position of the

anatomical target is from its actual position; (2) speed, i.e., how long does the algorithm take to acquire the results; and (3) robustness, i.e., how well does the algorithm deal with noise and outliers. The requirements for accuracy and robustness are equal for all stages of the procedure, where TREs within several millimeters can be considered as sufficient for most medical procedures. The speed of an algorithm is demanding in the intra-operative stage, and its requirement is from several seconds to a couple of minutes, while the time constraint can usually be relaxed for the other procedure stages [6].

The iterative closest point (ICP) method for the registration of 3D point clouds has been widely used in a variety of fields including medical images, because of its good accuracy and fast speed. It was proposed by Yang et al. [102] and Besl et al. [94]. In order to heighten registration accuracy between the extracted 3D fetal facial surface and a designed 3D fetal modal, an ICP-based method that is composed of coarse and fine registrations is proposed in this thesis according to the detected key facial features via 3D fetal face detection.

2.2.5 Visualization

The goal of data visualization for IGP is to briefly convey the relevant information for the successful completion of the intervention. Thus, the acquired medical images can be viewed by superimposing additional information such as tool locations or underlying anatomical structures. There are three ways to visualize tomographic images: surface rendering, volume re-slicing, and direct volume rendering (DVR). Surface rendering and DVR methods combined with segmentation are the only approaches for explicitly visualizing individual anatomical structures. Volume re-slicing and most DVR methods have no identification of anatomical structures, since visualization is done in 2D, and thus two or three orthogonal views are necessary to understand the underlying 3D anatomical structures.

2.2.6 Human Computer Interaction

The technologies associated with human computer interaction (HCI) can be divided into two categories: interaction techniques and information presentation. In the early stage, the standard keyboard and mouse were utilized as user input, and information was presented on standard computer monitors using the traditional axial,

sagittal, and coronal views. This approach may be acceptable for the pre and post -operative stages, but it is not suitable for intra-operative use. In the intra-operative stage, it is difficult for the physician to directly interact with the system [103]. Several technologies can allow the physician directly interact with the system, involving touch screens, trigger like input devices such as foot switches, tracked virtual keypads, speech recognition systems, and computer vision based gesture recognition techniques.

2.3 Conclusions

The research on medical image processing plays an important role on improving the performance of the IGP system. As the three main topics shown in Fig. 1.8 and explained in Section 1.3 and 1.4 have not yet been researched, several related technologies on medical image processing are studied and improved in this thesis for increasing location accuracy and efficiency of the system, which are discussed in the following chapters.

Chapter 3 Overview of the Proposed Methods

3.1 Prospective Surgical Navigation System

As outlined in Section 1.4 , based on fetal phantoms, the designed internal structure, real-time 3D US images, and a novel 3D US calibration system, the prospective surgical navigation system is designed to guide the FETO surgery by locating the surgical tool, the fetus and the critical internal organs during the operation. In addition, from a fetoscopic video sequence, fetal face and mouth are detected and tracked in real-time to provide an effective and timely feedback to the robot control system of the surgical tool.

3.1.1 Processes of the Prospective Surgical Navigation System

Figure 3.1 shows several steps of the prospective surgical navigation system, which consists of an offline and online process, where the fetus is supposed to be immobilized during the operation. The gray numbers ① and ② correspond to the following (1) and (2) in offline process, and the blue numbers ① through ⑥ correspond to the following (1) through (6) in online process. Note that this thesis uses resin fetal phantoms for offline and online processes because of difficulties in using real fetuses for the experiments.

(1) Offline Process

(1) 3D fetal modal with oral cavity and airways is reconstructed from real physical fetal phantoms (Section 7.2);

(2) Transformation matrix of the 3D US calibration system is estimated by the novel cones' phantom and an electromagnetic (EM) tracking device (Section 4.3);

(2) Online Process

(1) 3D fetal US images are obtained by scanning a fetal phantom in a water tank with a Philips V6-2 3D US probe (Section 5.2.2);

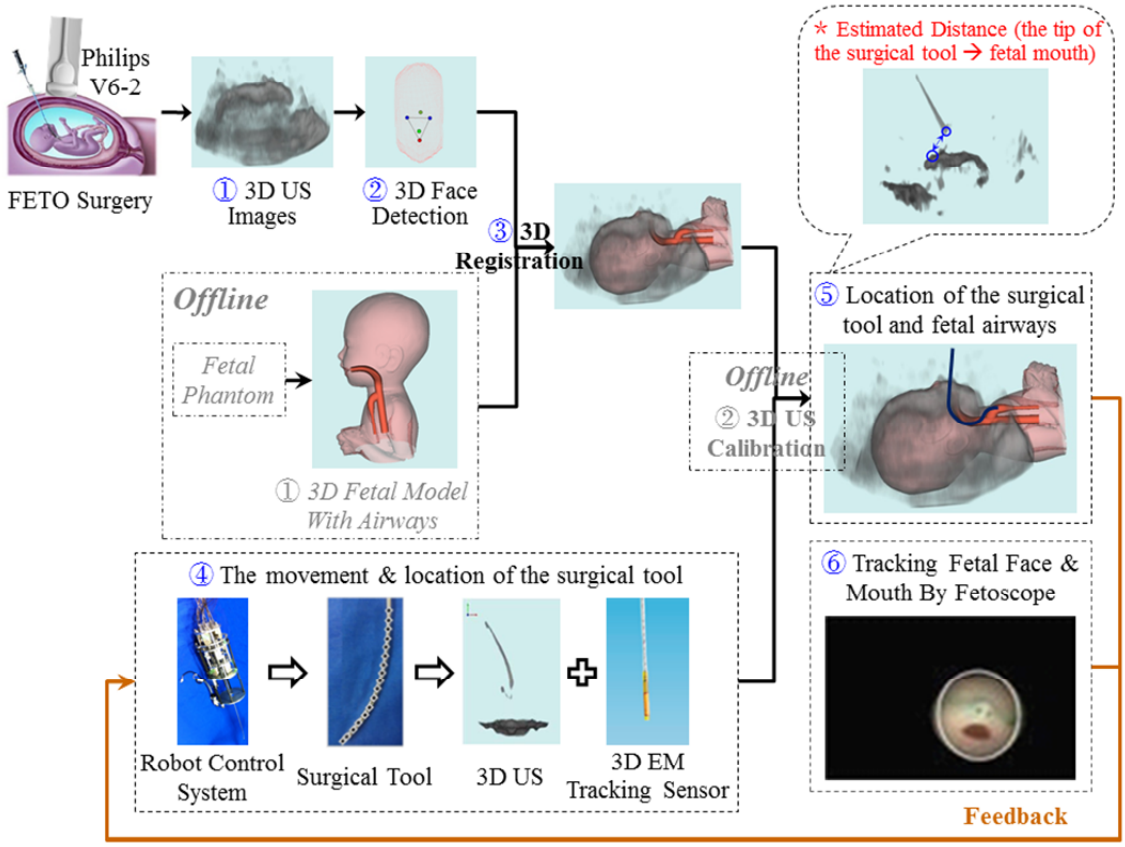


Figure 3.1 Process of the prospective surgical navigation system.

(2) 3D fetal facial surface is extracted from the obtained 3D US images, and key feature points (i.e., nose tip, two eyes' inner corners, nose upper bridge and mouth) of the extracted 3D fetal face are detected (Section 5.2.2);

(3) The 3D fetal model with the structure of internal organs is registered with the extracted 3D fetal face with the detected key feature points, and located in the 3D US space (the coordinate system of 3D US images) (Section 7.3).

(4) The surgical tool is detected and extracted from the 3D US images, and located in the 3D US space;

(5) The 3D fetal model with oral cavity and airways, the extracted 3D fetal facial surface from 3D US data, and the surgical tool are accurately located in the 3D EM tracking space (the world coordinate system of the 3D EM transmitter), via the transformation matrix of the 3D US calibration system estimated in the offline process; where, in the future, the distance between the tip of the surgical tool and the detected fetal mouth will be estimated before the surgical tool is inserted into the fetal mouth, and utilized as a feedback to the robot control system of the surgical tool;

(6) To guide the surgical tool into the fetal mouth, the fetal mouth are detected and tracked in real-time from a fetoscopic video sequence, so that the system can provide an effective and timely feedback to the robot control system of the surgical tool for guiding movement of the surgical tool in the future.

3.1.2 Three Main Topics

The three main topics of this thesis are contained in the online process, whose details of each topic are explained as follows:

(1) 3D Fetal Face Detection

Several steps are included in the 3D fetal face detection, as shown in Fig. 3.2, and are detailed in Chapter 5 .

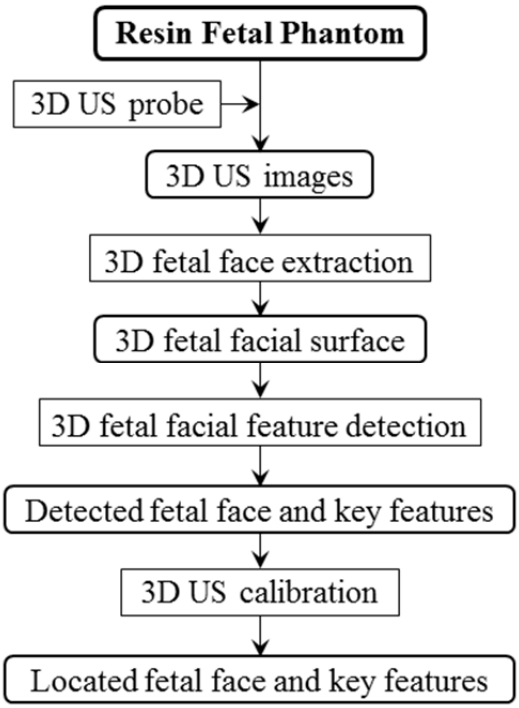


Figure 3.2 Process of 3D fetal face detection from 3D US images.

Step-1: 3D US images are obtained by scanning a resin fetal phantom in a water tank using a 3D US probe;

Step-2: 3D fetal facial surface is extracted from 3D US images;

Step-3: 3D fetal face and key facial features are detected by a boosting traversal scheme based on the spatial relations between each facial feature;

Step-4: The detected 3D fetal face and key facial features are located in 3D tracking space based on the 3D US calibration system.

(2) Surgical Tool Location and Fetal Mouth Tracking

For surgical tool location, it can be realized by the following steps, as shown in Fig. 3.3. The methods of the surgical tool location are described in Chapter 6 .

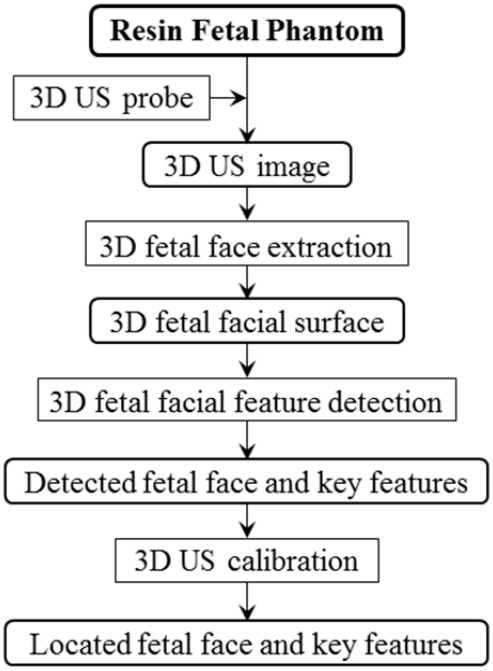


Figure 3.3 Process of surgical tool location from 3D US images.

Step-1: 3D US images are obtained by scanning a bending surgical tool in a water tank using a 3D US probe;

Step-2: The 3D skeletons of the surgical tool are extracted from 3D US images;

Step-3: The extracted 3D skeletons of the surgical tool are located in 3D tracking space based on the 3D US calibration system.

For fetal mouth tracking, it can be realized by the steps in Fig. 3.4, which will be illustrated in detail in Chapter 6 .

Step-1: Fetoscopic images are recorded by a fetoscopy when it moves with the

movement of the surgical tool;

Step-2: Fetal face is detected by a classifier from fetoscopic images;

Step-3: Fetal face is tracked by Camshift algorithm, and a region of interest (ROI) is given after tracking;

Step-4: Fetal mouth is detected and tracked in the ROI.

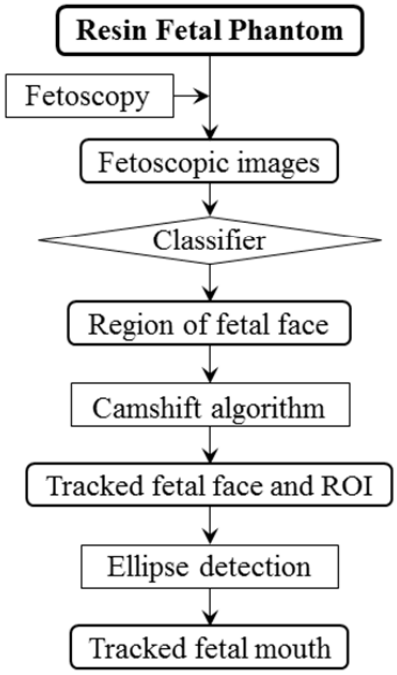


Figure 3.4 Process of fetal mouth tracking.

(3) Trachea Location

As a critical internal organ for the FETO surgery, fetal trachea can be located by the steps in Fig. 3.5, which will be explored in Chapter 7 .

Step-1: 3D US images are obtained by scanning a resin fetal phantom in a water tank using a 3D US probe;

Step-2: 3D fetal facial surface is extracted from 3D US images;

Step-3: 3D fetal face and key facial features are detected;

Step-4: The 3D fetal model reconstructed in the offline process is registered with the extracted 3D fetal facial surface to be located in 3D US space;

Step-5: Position of fetal trachea is estimated in 3D tracking space based on the 3D fetal model located in 3D US space and the 3D US calibration system.

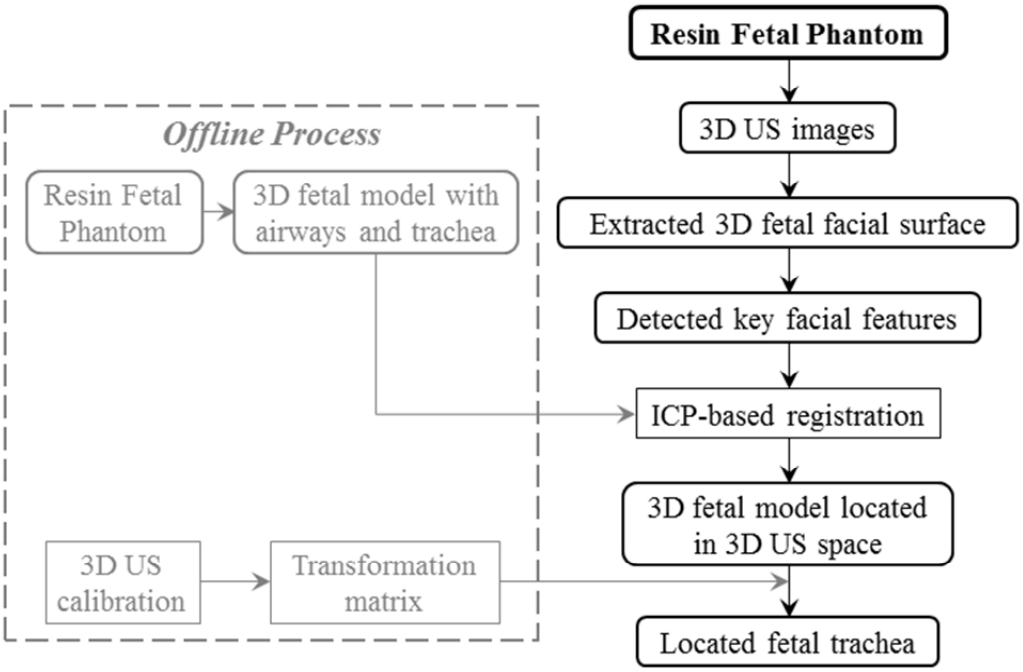


Figure 3.5 Process of fetal trachea location from 3D US images.

Chapter 4 Basic Technologies for Three Main Topics

Before discussing the three main topics of this thesis, three basic technologies on image segmentation, ellipse detection, and 3D US calibration are discussed in this chapter, and the relationship between these technologies and the three main topics that will be represented in Chapter 5, 6 and 7 has been portrayed in Fig. 1.9.

(1) An Improved Kernel-based Fuzzy C-Means Algorithm

An improved Kernel-based Fuzzy C-means Algorithm (iKFCM) is proposed to reduce the effect of noise for medical image segmentation, which applied k-nearest neighbor model and a neighborhood controlling factor by estimating image contextual constraints to optimize the objective function of conventional KFCM method. Conventional KFCM algorithms classify each pixel in image only by its own gray value, but the proposed method classifies by the gray values of its neighborhood system. For this reason, the proposed iKFCM algorithm has a strong robustness for image noise in image segmentation.

The algorithm is applied for segmentation of 2D US images (Chapter 5).

(2) An Improved Iterative Randomized Hough Transform Algorithm

An improved IRHT algorithm is proposed to detect partial ellipse in medical images, especially for ultrasound images. This work is different from previous ones in several aspects. First, the pre-processing guarantees the obtained pixels have a high possibility of generating one elliptic hypothesis with the selected edge. Second, the quasi random scheme reduces the requirement on the number of samples, i.e., a small number of samples can promise a robust and accurate result, which improves the efficiency of the algorithm. Third, the least square approach for parameter estimation, and the number (N) of pixels on the ellipse for updating the ROI heighten the accuracy and robustness of the algorithm.

This algorithm is applied for detection of fetal head in 2D ultrasound images (Chapter 5), fetal mouth in fetoscopic images (Chapter 6), and position of each cone in the 3D US calibration system (Section 4.3).

(3) A 3D Ultrasound Calibration System

3D ultrasound calibration aims to determine the spatial transformation for mapping points from the 3D US image coordinate system to the world coordinate system in 3D tracking space. The transformation comprises six parameters - 3 translations in the direction of the x, y, and z axes and the 3 rotations, azimuth, elevation, and roll, about these axes. In this thesis, a novel point-based phantom consisting of 12 resin cones is proposed for 3D US calibration. It is easy to construct and scan the phantom, and only simple experimental setup is required.

This system is applied for location of 3D key fetal facial features detected from 3D US images (Chapter 5), surgical tool location from 3D US images (Chapter 6), and estimating the position of the fetal trachea (Chapter 7).

4.1 An Improved Kernel-based Fuzzy C-Means Algorithm

4.1.1 Introduction

Medical image segmentation is an important step for medical image processing, which is a complex and challenging task, especially in brain magnetic resonance (MR) image segmentation. The purpose of medical image segmentation is to partition an image into several non-overlapping regions, which are homogeneous with respect to some characteristics such as grayscale or texture.

A large number of approaches are available for MR image segmentation [104-106]. However, for MR images, there are many uncertain factors, such as noise, bias field and partial volume effect, which make MR images essentially fuzzy [107]. Therefore, fuzzy clustering methods have been more widely used in MR image segmentation.

Fuzzy c-means (FCM) algorithm [60] is a typical clustering algorithm, which uses Euclidean distance to measure similarity among image pixels, but it is just effective in clustering ‘spherical’ data. In order to cluster more general dataset, a kernel-based fuzzy c-means (KFCM) algorithm [61, 62] is proposed by exploiting kernel function to measure data’s similarities instead of Euclidean distance. However, KFCM method is easily affected by noise owing to ignoring spatial information in images.

Recently, numerous methods [63-68] have been proposed by exploiting image spatial correlation to advance segmentation capability for low signal-to-noise ratio (SNR) images. In an image, the pixels should be highly relevant, i.e. the pixels in their immediate neighborhoods should possess almost the same characteristics. That is why, the spatial constraints among neighboring pixels is an important feature that can be contributing to image segmentation.

Herein, an improved KFCM method (iKFCM) [69] is proposed by cooperating with spatial information to optimize the objective function in conventional KFCM method so as to reduce the effect of noise. In conventional KFCM algorithms, the classification of every pixel in images is only determined by its own gray value. However, in the proposed iKFCM method, the classification of every pixel is also determined by the gray values of its neighbourhood system. For this reason, the proposed iKFCM method has a strong robustness for image noise in image segmentation. In experiments, some synthetic grayscale images and simulated brain MR images are used to assess the performance of iKFCM in comparison with other fuzzy clustering methods.

4.1.2 Algorithms

The Kernel-based Fuzzy C-means Algorithm (KFCM) [61] can perform a nonlinear mapping to a high dimensional feature space via a kernel function to replace Euclidean distance in FCM algorithm.

(1) Conventional KFCM Algorithm

The grayscale dataset of 2D objective image is given as $X = \{x_i, i = 1, \cdots, N\} \subset R^s$ in an s-dimensional space R^s, where x_i is the gray value of pixel i, N is the number of image pixels. The KFCM algorithm classifies X into L fuzzy subsets by minimizing the following objective function

$$J_m(U,V) = \sum_{k=1}^{L}\sum_{i=1}^{N} u_{ki}^m \| \Phi(x_i) - \Phi(v_k) \|^2 \tag{4.1}$$

where $\Phi(x)$ is a nonlinear map function defined as $\Phi: x \to \Phi(x) \in F, x \in X$. F is a transformed feature space with a higher dimension. L is the number of clusters (so the

objective image will be partitioned into L clusters). N is the number of image pixels. u_{ki} is fuzzy membership of pixel i in cluster k, satisfying $0 \le u_{ki} \le 1$, and $\sum_{k=1}^{L} u_{ki} = 1$. $v_k, k = 1, \cdots, L$ denote the average gray value in cluster k. m is a parameter for controlling clustering fuzziness, generally $m = 2$.

Also, $||\Phi(x_i) - \Phi(v_k)||^2 = K(x_i, x_i) + K(v_k, v_k) - 2K(x_i, v_k)$, and $K(x, y) = \Phi(x)^T \Phi(y)$ is an inner product kernel function. Here, Gaussian kernel function $K(x, y) = \exp(\frac{-||x-y||^2}{\sigma^2})$ is exploited, then $K(x, x) = 1$. Thus, the objective function is revised as :

$$J_m(U,V) = 2\sum_{k=1}^{L}\sum_{i=1}^{N} u_{ki}^m (1 - K(x_i, v_k)) \tag{4.2}$$

In Gaussian kernel function, σ is width of this function. Since σ is described as a dispersion, σ^2 is estimated with the following formula through the sample variance,

$$\sigma^2 = \sum_{i=1}^{N} || x_i - \overline{x} ||, \quad and \quad \overline{x} = \sum_{i=1}^{N} x_i / N \tag{4.3}$$

Thus, the update equations for minimizing $J_m(U, V)$ with the necessary conditions are as follows:

$$u_{ki} = \frac{(1 - K(x_i, v_k))^{-1/(m-1)}}{\sum_{j=1}^{L} (1 - K(x_i, v_j))^{-1/(m-1)}} \tag{4.4}$$

$$v_k = \frac{\sum_{i=1}^{N} u_{ki}^m K(x_i, v_k) x_i}{\sum_{i=1}^{N} u_{ki}^m K(x_i, v_k)} \tag{4.5}$$

An initial arbitrary value is adopted for each cluster center. Through iterative processing, the KFCM algorithm converges by comparing the changes in the membership function or the clustering centers between two successive iteration steps.

(2) Improved KFCM Algorithm

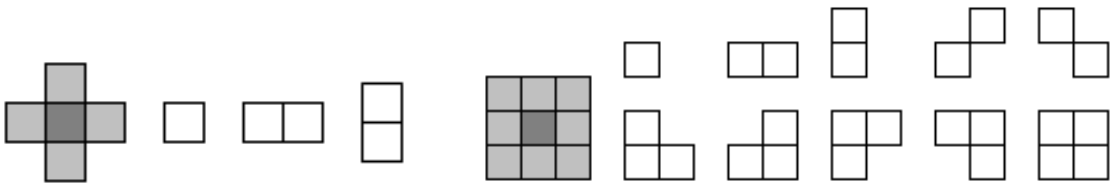
The application of contextual constraints is essential for a capable image analysis

system. Also spatial information among pixels indicates that neighboring pixels possess uniform feature values, and the probability that they belong to the same cluster is large. This spatial relationship is beneficial for suppressing influence of image noises, but conventional KFCM algorithm did not utilize the spatial relationship. Though the improved FCM method (here, it is called as sFCM) in [65] utilized the spatial relationship to modify the fuzzy membership u_{ki} by a spatial membership function, to some extent that will weaken the role of neighbourhood system. Besides, the method proposed by G.C. Karmakar and L.S. Dooley in [108] gives a better description for spatial relationship, which indicates every pixel can be endued with a fuzzy set. Further combined with *k*-nearest neighbor model, for pixel *i*, its membership function with spatial information is defined as:

$$p_{ki} = \frac{n_k}{\sum_{k=1}^{L} n_k} \tag{4.6}$$

where, p_{ki} denotes the membership of pixel *i* in cluster *k*, n_k is the number of the pixels in pixel *i*'s neighbourhood system N_i belonging to cluster *k* (the initial classification of every pixel in N_i is achieved by conventional KFCM algorithm). The general neighbourhood systems based on Euclidean distance are shown in Fig. 4.1. In experiments, a comparison of using first order neighborhood system and second order neighborhood system is given for the proposed method.

If the number of pixels in N_i belonging to cluster *k* is large, the probability of pixel *i* in cluster *k* should be large. Therefore, if the pixels in N_i are all classified into the same cluster, the membership of pixel *i* in the same cluster should be 1.



(a) (b)

Figure 4.1 Neighborhood systems and their cliques. (a) A first order neighborhood system with its cliques, (b) a second order neighborhood system with its cliques.

Then, p_{ki} is applied with spatial information and a neighborhood controlling factor α to modify the membership degree matrix of conventional KFCM algorithm.

Specially, fuzzy membership u_{ki} is converted into $u_{ki} = (1-\alpha+\alpha\cdot p_{ki})u_{ki}$, where α is constant for controlling the effect of neighborhood system, increased with noise intensity, and ranges $0 < \alpha < 1$. Accordingly, the new objective function is revised as:

$$J_m(U,V) = 2\sum_{k=1}^{L}\sum_{i=1}^{N}((1-\alpha+\alpha\cdot p_{ki})u_{ki})^m(1-K(x_i,v_k)) \tag{4.7}$$

According to the constraint of u_{ki} and Lagrange multipliers, the new iterative equations are derived as:

$$u_{ki} = \frac{(1-\alpha+\alpha\cdot p_{ki})(1-K(x_i,v_k))^{-1/(m-1)}}{\sum_{j=1}^{L}(1-K(x_i,v_j))^{-1/(m-1)}} \tag{4.8}$$

$$v_k = \frac{\sum_{i=1}^{N}u_{ki}^m K(x_i,v_k)x_i}{\sum_{i=1}^{N}u_{ki}^m K(x_i,v_k)} \tag{4.9}$$

The improved KFCM algorithm (iKFCM) is performed via the following steps.

Step-1 Set the number of cluster *L*, parameter $m = 2$, neighborhood controlling factor $\alpha = 0.5$, initial clustering centers v_k (via histogram statistics of image gray value), the smallest error $\varepsilon = 0.05$ and $p = 0$ (p is a counter). Next, get initial image segmentation results by conventional KFCM algorithm.

Step-2 Calculate probability p_{ki} by Eq. (4.6).

Step-3 Update fuzzy membership u_{ki} and clustering centers v_k by Eq. (4.8) and (4.9), $p = p + 1$.

Step-4 If $max[|u_{ki}^{(p)} - u_{ki}^{(p-1)}] \le \varepsilon$, then goto Step-5, else return to Step-2.

Step-5 Acquire the final segmentation results which are determined by fuzzy membership of each pixel.

4.1.3 Results

In order to verify the validity of the new method, conventional KFCM algorithm,

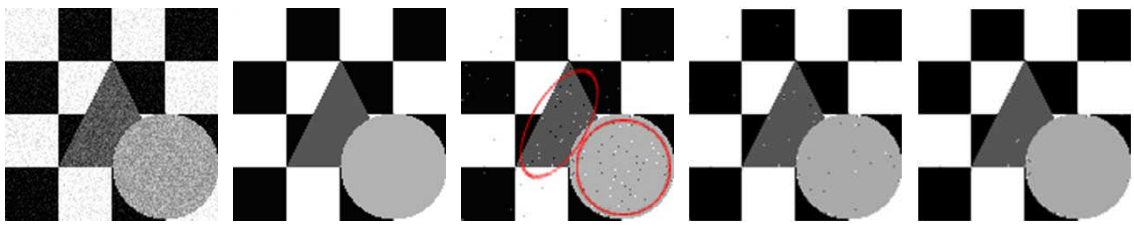
an improved FCM method with spatial information in [65] (here, it is called as sFCM) and the improved KFCM algorithm (iKFCM) are implemented in experiments. All algorithms are implemented under the same initial values and the same stopping condition with $\varepsilon = 0.05$. In brain MR image segmentation, the images are supposed to be segmented into four clusters: gray matter (GM), white matter (WM), cerebrospinal fluid (CSF) and background (BG). For quantitatively evaluating the ability of the above algorithms, misclassification rate (MCR) is defined as:

$$MCR = \frac{the\ No.of\ misclassified\ pixels}{the\ No.of\ all\ pixels} \times 100\% \tag{4.10}$$

where $MCR \in [0,1]$. Smaller MCR values correspond to better image segmentation performances. All experiments are conducted by VS 2003 in the PC of Intel® Core™2 CPU 6600 @ 2.40GHZ with 2GB memory.

(1) 24-bit synthetic gray images

24-bit synthetic grayscale images with 128×128 pixels are used, where their resolutions are 74.25 pixels per inch. The images are supposed to be classified into four clusters. Figure 4.2 gives a compassion of segmentation results of three approaches for a synthetic gray image with 7% Gaussian noise. The result in Fig. 4.2 (c) is calculated by conventional KFCM algorithm, the results in Fig. 4.2 (d) and Fig. 4.2 (e) are separately calculated by the improved method through applying first (iKFCM1) and second (iKFCM2) neighborhood system.



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

Figure 4.2 Comparison of segmentation results of three methods for a synthetic grayscale image with 7% Gaussian noise. (a) original image, (b) ground truth, (c) using conventional KFCM method, (d) using the improved KFCM method with first order neighborhood system ($iKFCM_1$), (e) using the improved KFCM method with second order neighborhood system ($iKFCM_2$).

According to the segmentation results in (c), (d) and (e) in Fig. 4.2, it can be

learned that there are the most misclassified points in the circles of Fig. 4.2 (c) in the result of conventional KFCM, which indicates the improved methods of using two different neighborhood systems achieve more accurate segmentation results. Also it can be known that the performance of iKFCM2 is superior to that of iKFCM1, which indicates the performance of the second-order neighborhood system is better than that of the first-order neighborhood system.

Table 4.1 *MCR*s (%) of three methods for synthetic images with different levels of Gaussian noise.

Noise level (%)	**7%**	**9%**	**11%**	**13%**
KFCM	0.85	3.72	8.43	13.56
iKFCM$_1$	0.25	1.78	5.13	8.86
iKFCM$_2$	**0.13**	**1.25**	**4.37**	**8.04**

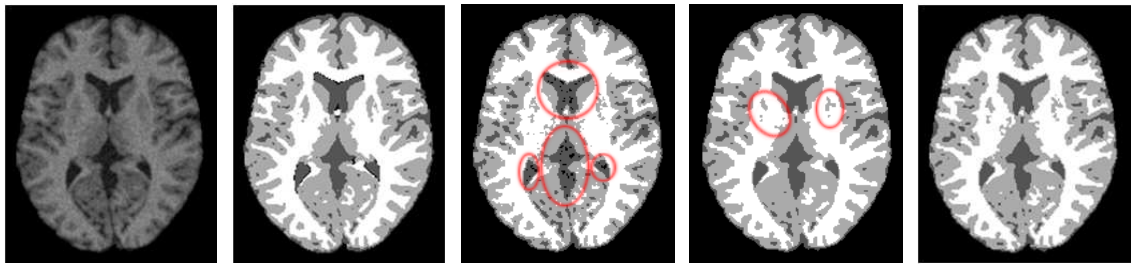
Table 4.1 shows *MCR*s of three methods for synthetic images with different levels of Gaussian noise. As shown in Table 4.1, the *MCR*s of conventional KFCM are the largest than those of the improved methods applying two neighborhood systems. Also the data in Table 4.1 reveal that the performance of iKFCM2 using the second neighborhood system is better than that of iKFCM1 using the first neighborhood system. In conclusion, iKFCM2 has the best segmentation performance in this test. Therefore, the second neighborhood system is used for the following experiments.

(2) Simulated Brain MR Images

In this section, simulated brain MR images in Brainweb (http://www.bic.mni.mcgill.ca/brainweb/), which are called gold standard of image segmentation, are used for validation. Each data set is composed of 258×258 pixels, thickness of layer is *1 mm*, T_1 weighted. The noise in the simulated images has Rayleigh statistics in the background and Rician statistics in the signal regions. The "percent noise" number represents the percent ratio of the standard deviation of the white Gaussian noise versus the signal for a reference tissue. Several simulated data sets (T_1, voxel size: $1\,mm^3$) with variations of the parameters “noise (n)” (ranging from 1 to 9%) are chosen. This selection covers the whole range of the parameter values available in BrainWeb so that the comparability with real data can be considered as sufficient to test the robustness of the different methods at varying image qualities [109].

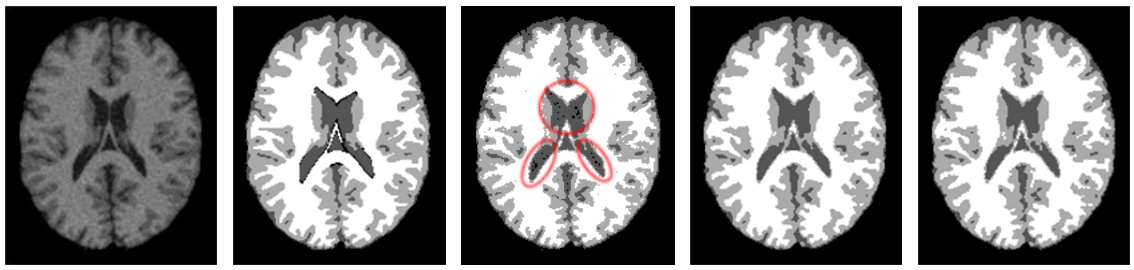
(2-1) Transverse Slices

In the first part, three groups of transverse slices with several levels of noise in $Z = 8\ mm$, $Z = 18\ mm$ and $Z = 28\ mm$ are used for testing several methods in brain image segmentation, where Z is a axial coordinate in a three-dimension space.



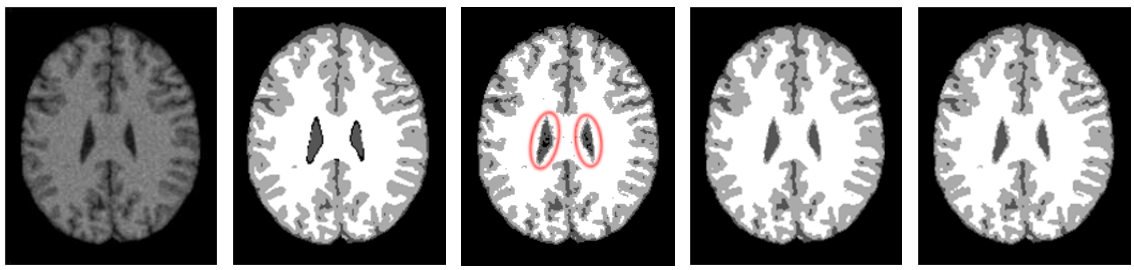
(a) (b) (c) (d) (e)

Figure 4.3 Comparison of segmentation results of three methods for a simulated brain MR transverse slice in $Z = 8mm$ with 9% Gaussian noise. (a) original image, (b) ground truth, (c) using conventional KFCM method, (d) using sFCM, (e) using iKFCM.



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

Figure 4.4 Comparison of segmentation results of three methods for a simulated brain MR transverse slice in $Z = 18mm$ with 9% Gaussian noise. (a) original image, (b) ground truth, (c) using conventional KFCM method, (d) using sFCM, (e) using iKFCM.



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

Figure 4.5 Comparison of segmentation results of three methods for a simulated brain MR transverse slice in $Z = 28mm$ with 9% Gaussian noise. (a) original image, (b) ground truth, (c) using conventional KFCM method, (d) using sFCM, (e) using iKFCM.

Figure 4.3, Figure 4.4 and Figure 4.5 show segmentation results by the three methods, respectively, for simulated brain MR transverse slices in $Z = 8$ *mm*, $Z = 18$ *mm* and $Z = 28$ *mm* adding 9% Gaussian noise. sFCM used in spatial information in [65], which utilizes the second order neighborhood system for the purpose of being consistent with the iKFCM method. There are many more noise dots in circles of (c) of each figure than in the same areas in (d) and (e) of each figure, which indicates conventional KFCM method is easily affected by image noise owing to be deficient in image spatial information. Also there are some over-segmentations in some region edges for sFCM method, especially in circles of Fig. 4.3 (d), compared with the same areas in Fig. 4.3 (e). Accordingly, all experimental results demonstrate that the improved approach (iKFCM) achieves the best result in this processing. Moreover, the computational efficiency of the iKFCM method is the same as that of conventional KFCM method, but far better than that of the sFCM method.

Table 4.2 presents a comparison of *MCR*s of three methods for simulated brain MR slices in various Z positions with different levels of Gaussian noise. The data in Table 4.2 denote that the *MCR*s of the iKFCM method for all simulated brain MR slices are the least than those of conventional KFCM and sFCM method. This means the proposed method is robust for image noise in the segmentation of different brain images. Furthermore, the stronger the intensity of noise in image, the better the segmentation property of the proposed method compared with conventional KFCM.

Table 4.2 *MCR*s (%) of three methods for simulated brain MR transverse slices with different levels of Gaussian noise.

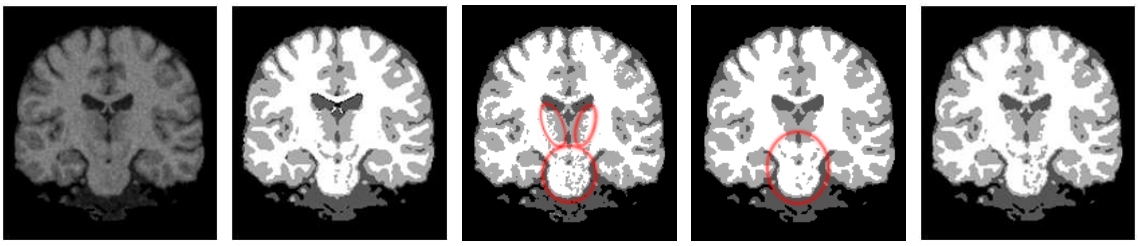
Slice Coords	***Z* = 8 *mm***				***Z* = 18 *mm***			
Noise Level (%)	**3%**	**5%**	**7%**	**9%**	**3%**	**5%**	**7%**	**9%**
KFCM	6.88	8.66	10.01	11.67	5.84	6.93	8.00	9.82
sFCM	6.32	6.78	7.65	8.16	5.72	6.41	7.04	7.83
iKFCM	**4.72**	**6.10**	**6.93**	**7.60**	**5.01**	**5.82**	**6.35**	**7.41**

Slice Coords	***Z* = 28 *mm***			
Noise Level (%)	**3%**	**5%**	**7%**	**9%**
KFCM	7.25	8.47	9.35	10.99
sFCM	5.92	6.13	6.82	7.31
iKFCM	**4.62**	**5.20**	**6.02**	**6.66**

(2-2) Coronal Slices

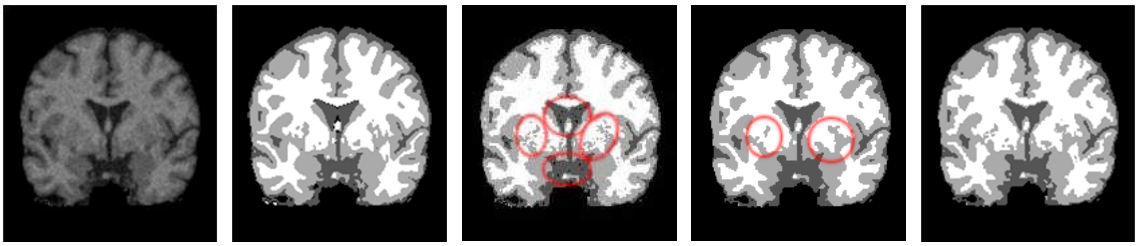
In this section, two groups of coronal slices in *Y = -18 mm* and *Y = 0 mm* are used for testing several methods in brain image segmentation, where *Y* is a coronal coordinate in a three-dimensional space.

Figure 4.6 and Figure 4.7 are the segmentation results of the three methods for simulated brain MR coronal slices in *Y = -18 mm* and *Y = 0 mm* adding 9% Gaussian noise. From the above two figures, the segmentation results in circles of (c) are heavily affected by image noise for conventional KFCM method. Some over-segmentation results have been trapped in circles of (d) by sFCM method, compared with the same areas in (e). Consequently, the iKFCM method not only exploits spatial information among image pixels to avert the effect of noise, but also effectively reduces over-segmentation induced by the complexity of medical images.



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

Figure 4.6 Comparison of segmentation results of three methods for a simulated brain MR coronal slice in *Y = -18 mm* with 9% Gaussian noise. (a) original image, (b) ground truth, (c) using conventional KFCM method, (d) using sFCM, (e) using iKFCM.



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

Figure 4.7 Comparison of segmentation results of three methods for a simulated brain MR coronal slice in *Y = 0 mm* with 9% Gaussian noise. (a) original image, (b) ground truth, (c) using conventional KFCM method, (d) using sFCM, (e) using iKFCM.

In Table 4.3, a comparison of *MCR*s of the three methods for simulated brain MR slices in various *Y* positions with different levels of Gaussian noise, it can be known that

though the improvement in *MCR*s of the iKFCM methods changes with varied *Y* positions and varied noise levels, the overall effect is the best among these three methods in this comparison.

Table 4.3 *MCR*s (%) of three methods for simulated brain MR coronal slices with different levels of Gaussian noise.

Slice Coords	***Y = -18 mm***			
Noise Level (%)	**3%**	**5%**	**7%**	**9%**
KFCM	4.79	5.92	7.19	8.15
sFCM	4.96	5.43	5.89	6.46
iKFCM	**4.23**	**4.68**	**5.65**	**6.19**

Slice Coords	***Y = 0 mm***			
Noise Level (%)	**3%**	**5%**	**7%**	**9%**
KFCM	4.50	5.31	6.46	7.57
sFCM	4.85	5.00	5.50	6.19
iKFCM	**3.87**	**4.52**	**4.94**	**5.85**

4.1.4 Conclusions

In this section, an improved KFCM method (iKFCM) with spatial information has proposed to modify the objective function of conventional KFCM method in order to reduce the effect of noise. In the proposed approach, the classification of every pixel in images is not only determined by its own gray value, but also determined by the gray values of its neighborhood system. The experimental results for the segmentation of synthetic gray-scale images and simulated brain MR images, show that the proposed method has a strong robustness for image noise, and also achieves a better segmentation performance than other fuzzy clustering methods.

In addition to the application in segmentation of brain MR images, the proposed iKFCM method is also applied for segmentation of 2D US images in this thesis, where bright parts of fetal skulls are divided from 2D US images for 3D fetal facial surface extraction, which is described in Chapter 5 .

4.2 An Improved Iterative Randomized Hough Transform Algorithm

Ellipse is one of the most commonly occurring geometric shapes in real images. Thus, detection of ellipses in a robust, reliable and efficient manner from real images provides a good image analysis tool for pattern recognition.

In medical image processing, ellipse detection is also an important means for many clinical applications, such as fetal head detection on ultrasound (US) images for fetal growth evaluation [110] and fetal weight estimation [111], and fetal mouth detection from a fetoscopic video sequence for a surgical guidance [112], etc. However, medical images, such as US images and fetoscopic images are inevitably accompanied by moderate noise, curve discontinuity, shape deformation and image artifacts, which make the detection of geometric shapes like ellipses very difficult in medical images and thus technically interesting and challenging.

4.2.1 Introduction

(1) Literature Review

Ellipse detection in images has been an open research problem since long time ago. Initially, approaches like least squares fitting and Hough transform (HT) [72] were the main approaches for this purpose. These approaches generally use the mathematical model of ellipse and the edge pixels for detecting the ellipses. Least squares based methods [70, 71] usually cast the ellipse fitting problem into a constrained matrix equation in which the solution should minimize the residue in the mathematical model. The choice of constraints and solutions for constrained matrix equations has an impact on the performance and selectivity of the ellipse detection methods.

Hough transform (HT) and its variations seek an estimation of parameters by clustering in the parameter space [113, 114]. HT based methods are usually more robust than least squares based methods when the edge contours are not smooth as they use pixels for detecting ellipses instead of edge contours. However, HT based methods have two main problems. First problem is that HT is intensive in terms of computation and memory consumption because it uses a five dimensional parameters space. Second

problem is that since the pixels used in HT need not belong to the same edge contour, number of samples required for detecting each ellipse is very high.

To overcome some limitations of the HT, combinatorial Hough transform (CHT) [73], randomized Hough transform (RHT) [74], probabilistic Hough transform (PHT) [75], and dynamic generalized Hough transform (DGHT) [76] were proposed to improve the performance of HT for non-linear problems like ellipse detection. The PHT randomly selects a small subset of pixels for further processing. The other transforms share the following fundamental idea and are hereinafter referred to as RHTs. For an n-dimensional curve, a sample of n randomly picked pixels is used to solve for the n parameters of the curve. The count for the solution (an instance of the model) is increased by one in the parameter space. This process is repeated until enough solutions are accumulated so that there is a meaningful count distribution in the parameter space. The peak count and its coordinates in the parameter space correspond, respectively, to the most prominent curve instance in the image space and curve parameters. This random sampling (of n pixels) and convergent mapping (of n pixels to one solution) are the key mechanisms of RHTs, accounting for their favorable properties of small storage, high speed, good accuracy and reduced effects of correlated noise [74].

Random Sample Consensus (RANSAC) [77] algorithm shared the two key mechanisms of RHTs. The RANSAC also randomly samples n pixels to get an instance of the model. Instead of accumulating the counts as in the RHTs, the RANSAC checks the number of the consensus pixels that are within some error tolerance of the instance. If this number is greater than some threshold, a model is solved with the consensus pixels; otherwise, another set of n pixels is sampled and the process is repeated.

However, both RHTs and RANSAC may fail when strong noise can corrupt the curve-related peaks in the parameter space, or can generate a larger consensus set for a false model instance. A novel method named iterative randomized Hough transform (IRHT) [115, 116] was proposed for detection of partial ellipses under strong noise conditions. The IRHT iteratively applies a randomized Hough transform to a region of interest (ROI) in the image space. The ROI is updated based on the latest estimates of parameters. First limitation of the IRHT is that all extracted edge pixels that cannot be grouped for one elliptic hypothesis are used for ellipse detection, and second limitation is random selection of pixels for parameter estimation, which both limitations require a large number of samples and result in a low detection efficiency. Third limitation is that

the robustness of the algorithm is not strong, because it does not consider the number of pixels on the ellipse.

(2) The Proposed Method

In order to overcome the limitations of the original IRHT method, a pre-processing is given to remove redundant pixels and exclude edges that cannot generate one elliptic hypothesis with the considered edge by edge grouping after edge extraction. During ellipse detection, a quasi random scheme selects three pixels as the centers of three windows, and fit all edge points in these three windows to an ellipse by the least square approach [117] (Fig. 4.8). Meanwhile, the number of pixels on the ellipses is considered to determine the result for each iteration by selecting the detected ellipse with the maximum number of pixels on the ellipses, which is chosen from the top-M peaks in the accumulators of the entire detected ellipse samples, where M is a specified number of peaks in the accumulators of the detected ellipses.

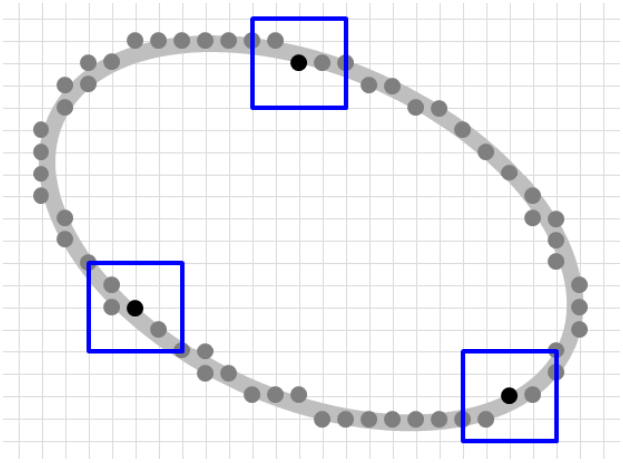


Figure 4.8 Quasi random scheme for a digital ellipse.

This work is different from previous ones in several aspects. First, the pre-processing guarantees the obtained pixels have a high possibility of generating one elliptic hypothesis with the selected edge. Second, the quasi random scheme reduces the requirement on the number of samples, i.e., a small number of samples can promise a robust and accurate result, which improves the efficiency of the algorithm. Third, the least square approach for parameter estimation, and the number of pixels on the ellipse for updating the ROI heighten the accuracy and robustness of the algorithm.

4.2.2 Methods

(1) Algorithm Overview

The flow diagram of the proposed algorithm is summarized in Fig. 4.9. The algorithm begins with edge extraction to detect skeletons (edges) of the areas in an image. In pre-processing, the edges that have a high probability of being on the same ellipse are obtained. Finally, the improved IRHT (iIRHT) algorithm [118, 119] is proposed to detect ellipse from the obtained edges.

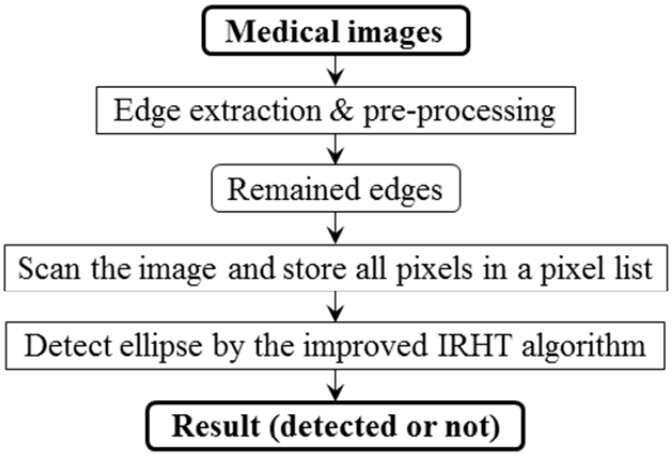


Figure 4.9 Flow diagram of the proposed algorithm.

(2) Edge Extraction

Let us take fetal head detection in US Images (Fig. 4.12 (a)) as an example, in which skeletons of fetal skulls are to be extracted as the brightest areas by edge extraction. First of all, a Gaussian filter with a 5×5 window is exploited to reduce the speckle noise (Fig. 4.12 (b)), and a white top-hat transform is operated to increase the contrast with a 9×9 structuring element (Fig. 4.12 (c)).

After that, the iKFCM algorithm [69] is applied to distinguish the brightest areas from the fetal US images (Fig. 4.12 (d)). In this way, the mean value μ_i and the standard deviation σ_i $(i = 0, \cdots, k-1)$ of each cluster can be calculated from the segmentation results. Since only main skeletons of fetal skulls are necessary as the input of the ellipse detection, it is not essential to extract the whole brightest areas comprising much noise.

Subsequently, a global thresholding is utilized to convert the intensity image into a binary image, and then the brightest areas are extracted from the background by a simple operation that compares image gray values with a threshold value *T* (Fig. 4.12 (e)). The following threshold value *T* verified in the experiments, is used to extract the brightest areas,

$$T = \mu_{k-1} - 0.75 \times \sigma_{k-1} \tag{4.11}$$

where, μ_{k-1} and σ_{k-1} are the mean value and standard deviation of the brightest areas, respectively (Here, the brightest areas are considered as the $k-1$ cluster).

Concerning binary image, a binary morphologic opening operation with a 2×2 structuring element is used to remove some small bright areas, and a closing operation with a 3×3 structuring element is used to smooth the boundaries of the large bright areas. In the end, the skeletons (or edge curves) of the brightest areas are extracted by distance transform [120] (Fig. 4.12 (f)).

(3) Pre-processing

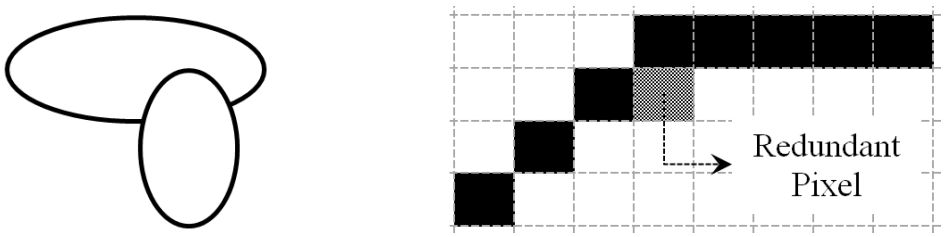
(3-1) Removal of Redundant Pixels

In Fig. 4.10 (b), if a junction on the extracted edges (Fig. 4.10 (a)) is zoomed, many redundant pixels are left around corners, which could lead to the errors in the ellipse detection stage. Those pixels present in the corners tend to form *L* shape, as shown in Fig. 4.11, and are eliminated as redundant pixels [121] (Fig. 4.12 (g)).

(3-2) Edge Grouping

After the removal of redundant pixels, edge curves are classified by the number of pixels on each connected curve. Those less than 10 pixel long are excluded because they may be the effect of noise or do not contribute to ellipse detection process. Meanwhile, the longest connected edge curve is selected as one part of the assumed target ellipse, and a search region, which is defined below, can be found such that the edge curves outside the search region of the selected edge curve can be certainly rejected from grouping with the considered edge curve. Such rejection is possible by defining a suitable search region such that the edge curves inside the search region are the only

edge curves that may be grouped with the considered edge curve for forming ellipses. Such a search region is shown in Fig. 4.13 (a). The shaded region R_2 is the search region and the edge curves that do not lie completely inside the search region are not grouped with the considered edge curve e_1 [122]. In Fig. 4.13, the shaded region shows the search region for the edge curve e_1. Based on the search region, it can be concluded that the edge curves e_2, e_3 and e_4 cannot be grouped with e_1 for generating one elliptic hypothesis. The search region can be found using the lines l_1, l_2 and l_3 and the midpoint of the edge curve P_{mid} shown in (b).



(a) (b)

Figure 4.10 Edges and redundant pixel.

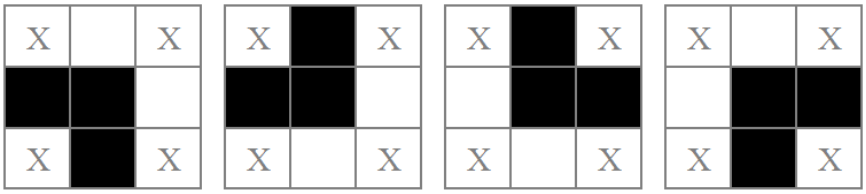


Figure 4.11 Redundant pixel to be removed.

- **Determining of the Search Region**

The search region *R* is defined as follows. For a given edge curve *e*, let the tangents to the edge curve at its end points P_1 and P_2 denote l_1 and l_2, respectively, and let the line segment connecting the end points P_1 and P_2 denote l_3. The tangents to the edge curves are computed using reference [123], and the value of the control parameter is chosen to be 4. The two tangents l_1 and l_2 and the line l_3 divide the image into two regions, R_1 and R_2, as shown in Fig. 4.13 (b). Then the search region R is determined as the region that does not contain P_{mid}, which is the middle point of the edge curve *e*. In Fig. 4.13 (b) and Fig. 4.12 (h), the search region *R* is determined as R_2, i.e., $R = R_2$, because $P_{mid} \notin R_2$.

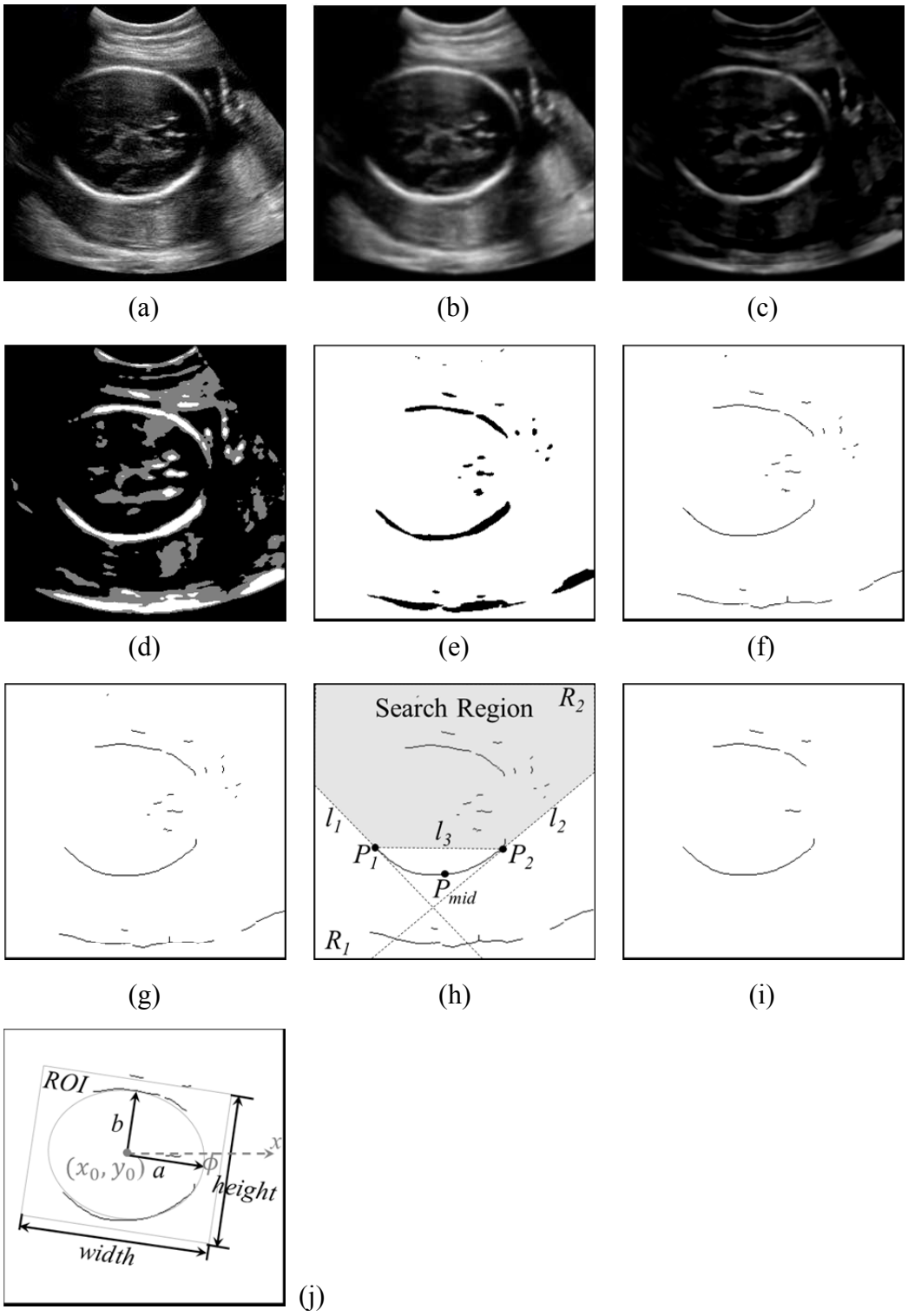


Figure 4.12 Edge extraction and pre-processing on an US image, (a) a clinical fetal US image, (b) the result of a Gaussian filter on (a) , (c) the result of a white top-hat transform on (b), (d) the segmentation result on (c), (e) the result of a global thresholding on (d), (f) the skeletons extracted from (e), (g) the result after removing redundant pixels, (h) the search region of the longest connected edge curve, (i) the reserved edges after edge grouping, (j) the ROI decided by the estimated ellipse.

- **Finding the Edge Curves within the Search Region of an Edge Curve**

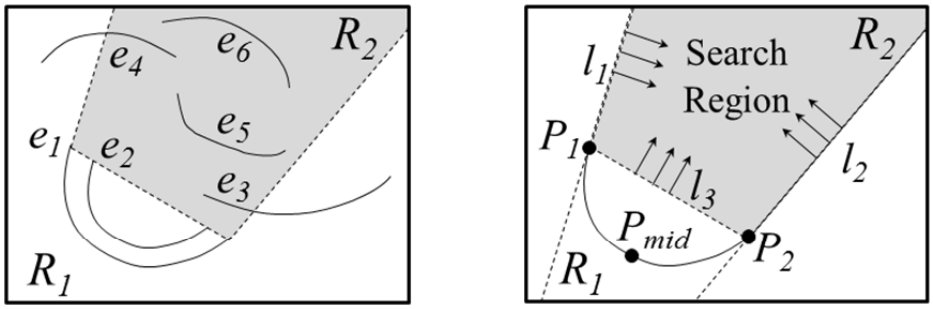
For a given edge curve *e*, after finding its search region *R*, the edge curves within the search region can be found as follows. An edge curve $e_i \in R$ if all the following three criteria $S1 - S3$ are satisfied:

① *Search region criterion 1 (S1)*: e_i and P_{md} are on the same side of l_1.

② *Search region criterion 2 (S2)*: e_i and P_{md} are on the same side of l_2.

③ *Search region criterion 3 (S3)*: e_i and P_{md} are on the opposite side of l_3.

After the edge extraction and pre-processing, the obtained curves have a high probability on the same ellipse (Fig. 4.12 (i)).



(a) (b)

Figure 4.13 Illustration of the detection of the search region, (a) illustration of the search region, and (b) definition of the search region.

(4) Iterative Randomized Hough Transform

In US images, fetal skulls often appear as the brightest areas with some discontinuities, because fetal skulls are not completely closed. Besides, some other structures also may generate bright spots in the image. Furthermore, various artifacts and noise are usually present in US images. Consequently, a useful head detection algorithm must effectively deal with these disturbances. The iterative randomized Hough transform (IRHT) [115] was recently developed for the detection of ellipse with large discontinuities and strong noise, derived from randomized Hough transform (RHT) [124] method. The following items provide a brief description of the RHT algorithm and the IRHT algorithm.

(4-1) Randomized Hough Transform (RHT)

In a binary image, the curve to be detected can be modeled by $f(c, z) = 0$, where

$c = [\alpha_1, \cdots, \alpha_n]^t$ comprises n-dimensional parameters, $z = (x, y)$ represents the coordinates of pixels on the curve. The RHT method first randomly takes a sample of n pixels, $z_i = (x_i, y_i), i = 1, \cdots, n$, and maps this sample into one point $c \in R^n$ in the n-D parameter space by solving a set of n-s equations $f(c, z_i) = 0$. If c is valid for ellipse, the counter at c is increased by one in the parameter space and stored in its corresponding accumulator. This process is repeated until a predefined number of valid samples (K) are processed. The location of the counter peak in the accumulators denotes a remarkable possibility of the curve in the image. For ellipse detection ($n = 5$), the following equation is suitable to be utilized [76, 125]:

$$x^2 + y^2 - U(x^2 - y^2) - V2xy - Rx - Sy - T = 0 \tag{4.12}$$

where, the five parameters, $[U, V, R, S, T]^t$, can be converted into the standard ellipse parameters $c = [x_0, y_0, a, b, \Phi]^t$, (x_0, y_0) are the center coordinates of the ellipse, a and b are its major and minor semi-axes, and Φ is the angle of rotation, then the ellipse eccentricity is given by $e = b/a$ and

$$U = \cos 2\phi \frac{1 - e^2}{1 + e^2} \tag{4.13}$$

$$V = \sin 2\phi \frac{1 - e^2}{1 + e^2} \tag{4.14}$$

$$R = 2x_0(1 - U) - 2y_0V \tag{4.15}$$

$$S = 2y_0(1 + U) - 2x_0V \tag{4.16}$$

$$T = \frac{2a^2b^2}{a^2 + b^2} - \frac{x_0R}{2} - \frac{y_0S}{2} \tag{4.17}$$

where, U and V depend only on Φ and e. In particular, for a circle, U and V are zero.

(4-2) Iterative Randomized Hough Transform

The IRHT method employs a randomized Hough transform (RHT) to a region of interest (ROI) in the image space by iterative parameter adjustment and reciprocal use of the image space and parameter space. In the RHT method, the ROI is always the whole image and it does not change during the whole process. However, to get more

accurate results, the ROI in the IRHT method is updated based on the latest estimates of parameters during the iteration process. Thus, noise pixels are gradually excluded from the ROI, and the estimation progressively becomes close to the target.

The steps of the IRHT algorithm are described in the following, and its flow diagram for ellipse detection is shown in Fig. 4.14.

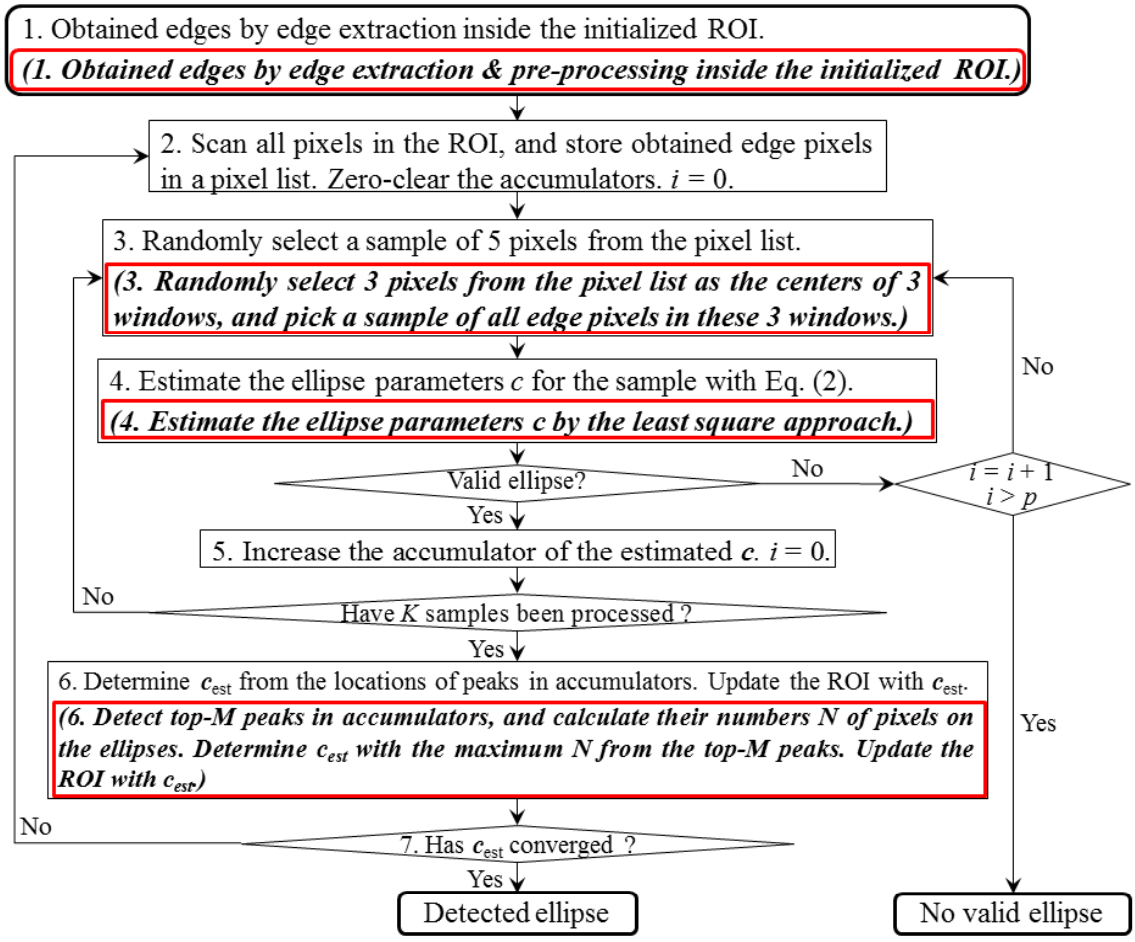


Figure 4.14 Diagrams of the IRHT and iIRHT algorithms (Black characters: IRHT; bold, italic characters surrounded by red rectangle: iIRHT; blocks with only black characters: common to IRHT and iIRHT).

Step-1 Edges are obtained by edge extraction, and the ROI is set as the whole image;

Step-2 Scan all pixels in the ROI, and store obtained edge pixels in a pixel list. Zero-clear the accumulators. Set $i = 0$, where i is a counter for false detections;

Step-3 A sample of 5 pixels is randomly selected from the pixel list;

Step-4 A parameter solution $c([x_0,\ y_0,\ a,\ b,\ \phi]^t)$ is estimated for this sample with Eq.

(4.12). To ensure that *c* represents a valid ellipse, the following conditions need to be met: x_0, y_0, a, and b are positive, x_0 and y_0 are not larger than the image width and height respectively, and *a* and *b* are not longer than one half of the diagonal length of the image. If *c* is not valid, then $i = i + 1$; subsequently, if $i \leq p$ (= a large number, such as 1000), then return to Step-3; otherwise, the detection process terminates and reports failure of detection;

Step-5 Increase the count for the accumulator that corresponds to *c*. If *K* samples have not been processed, then return to Step-3, where $K = 200$ is often used by the original IRHT method [115];

Step-6 A parameter estimate, c_{est}, is determined from the locations of the count peaks in the accumulators, and update the ROI based on c_{est} (see below for more specific procedure);

Step-7 If c_{est} has converged (see below for explanation of convergence conditions), output the result c_{est} detected in the final iteration as the final result; else, return to Step-2.

- **Update of the ROI**

A slightly larger region than the region of the estimated ellipse is used to compensate for uncertainties in the estimated parameter c_{est}. In Fig. 4.12 (j), the ROI is simply considered as a rectangle, whose center and angle of rotation are the same as (x_0, y_0) ((x_0, y_0) is indicated by a gray dot in Fig. 4.12 (j)) and Φ of the estimated ellipse, width equals $2a + \sigma_a$, and height equals $2b + \sigma_b$, where σ_a and σ_b indicate the standard deviations of *a*'s and *b*'s values from the found *K* samples.

- **Conditions for c_{est}'s Convergence**

If the ellipses detected in two successive iterations satisfy the following conditions, c_{est} is judged to be converged. The conditions are as follows: less than 2.5° in Φ; less than 2 pixels in each of x_0, y_0, a and *b*; and less than 6 pixels total in x_0, y_0, a and *b*.

(4-3) Improved Iterative Randomized Hough Transform

Based on the principle of the IRHT method, an improved IRHT (iIRHT) algorithm is proposed by introducing a quasi random scheme, the least square approach, and the number of pixels on the ellipse, to achieve high accuracy, robustness and efficiency of

the algorithm. The flow diagram of the iIRHT algorithm is similar as that of the original IRHT algorithm, but the improvements are applied to Step-1, Step-3, Step-4, and Step-6 as indicated by bold, italic characters surrounded by red rectangles in Fig. 4.14. Note that the other blocks and procedures are common to the original IRHT algorithm.

In Step-1, redundant pixels on edges are removed at first, and the edge pixels generating one elliptic hypothesis at a high possibility with the longest connected edge are reserved by edge grouping, and then the ROI is initialized as the whole image. In Step-3 (the quasi random scheme), 3 pixels are randomly selected from the pixel list as the centers of the three windows, where the distance of each selected pixel should be larger than 10. Then, a sample of pixels from the pixel list are selected in these 3 windows; In Step-4, all the selected pixels in the windows are fitted to an ellipse with a parameter solution $c([x_0,\ y_0,\ a,\ b,\ \phi]^t)$ by the least square approach. In Step-6 (the number of pixels on the detected ellipse), top-M peaks are detected from the accumulators, and their numbers N of pixels on the ellipses are calculated; Determine c_{est} with the maximum N from the top-M peaks, and update the ROI based on c_{est}. In addition, $K = 100$, which is smaller than that value ($K = 200$) of the original IRHT method [115], and contributes to a higher efficiency of the iIRHT algorithm.

4.2.3 Experiments

(1) Fetal Head Detection

Clinical fetal ultrasound images are provided by Nanjing Maternity and Child Health Care Hospital in China, each of which has had prior consent from the Mothers, before being used in this research. For the purpose of raising the performance of the algorithm, a priori information, reported by Hadlock et al. [126], is applied into this experiment, which shows that the eccentricity $e = b/a$ of the human fetal head has a mean (μ_e) of 0.783 and a standard deviation (σ_e) of 0.044. It could be used to construct a constraint as $\mu_e - 3\sigma_e \le e \le \mu_e + 3\sigma_e$, namely, $0.651 \le e \le 0.915$, and about 99.7% of fetal heads would have an eccentricity in this range. All programs of each method in C++ run on a quad-core 2.60 GHz computer with 8GB memory.

(1-1) One Clinical Ultrasound Image

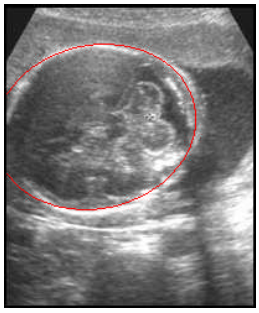
A clinical US image is applied to test the performance of several methods, and the

red ellipse in Fig. 4.15 (a) is detected by surgeons manually. The edges in Fig. 4.15 (b) are extracted by edge extraction, and the skeletons in Fig. 4.15 (c) are detected by pre-processing. For each method, 100 groups of the parameters of the detected ellipses are recorded in Fig. 4.15 (d)-(m) and Table 4.4.

Table 4.4 The means of the estimated parameters for several methods.

Methods	**Parameters**						
	x_0	y_0	a	b	$\Phi(°)$	N	*Time(s)*
RHT	73.43 ± 3.80	104.78 ± 3.15	83.27 ± 2.57	68.32 ± 2.10	92.39 ± 64.41	25.54 ± 15.01	3.07 ± 2.25
IRHT	70.40 ± 2.91	99.65 ± 2.44	79.63 ± 2.73	67.47 ± 1.28	160.21 ± 6.07	47.28 ± 27.20	2.60 ± 1.31
iIRHT (M=3)	73.31 ± 3.03	98.09 ± 1.18	81.03 ± 2.83	67.16 ± 1.43	161.06 ± 5.09	76.01 ± 18.91	0.87 ± 0.62
iIRHT (M=6)	**73.27 ± 2.13**	**98.14 ± 0.80**	**81.52 ± 2.20**	**66.83 ± 1.15**	**166.07 ± 4.48**	**85.18 ± 14.84**	**0.79 ± 0.43**
iIRHT (M=9)	73.37 ± 2.78	98.24 ± 0.88	81.40 ± 2.95	67.09 ± 1.04	165.01 ± 5.19	87.53 ± 16.14	0.84 ± 0.45

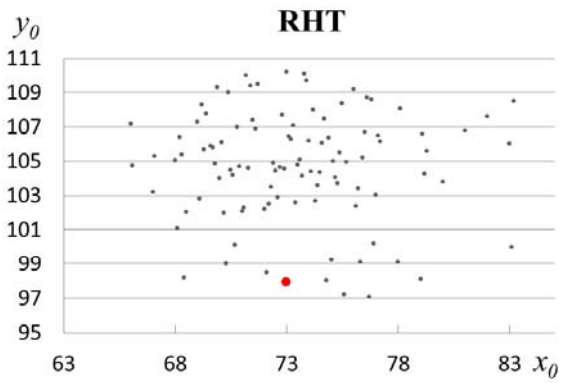
Figure 4.15 (d)-(h) enumerate the center coordinates of the detected ellipses via RHT, IRHT, iIRHT ($M = 3$), iIRHT ($M = 6$), and iIRHT ($M = 9$) algorithms, where the red points in each figure are the ground truth of the center coordinate of the ellipse, and the horizontal and vertical axes correspond to the x and y axes, respectively. It can be learnd that the detected center coordinates of RHT method are very divergent from the ground truth, those of IRHT method are relatively a little convergent from the ground truth, and those of iIRHT methods are very convergent from the ground, where iIRHT ($M = 6$) method performs best and acquires the most convergent results. On the other hand, based on the comparison of the means of each parameter detected via several methods also prove that the iIRHT ($M = 6$) method achieves the best accuracy of the estimated parameters, the best robustness and the best efficiency (0.79 ± 0.43 s) of the algorithm. Thus, in the following experiments and applications, M is selected as 6 for the proposed iIRHT algorithm.



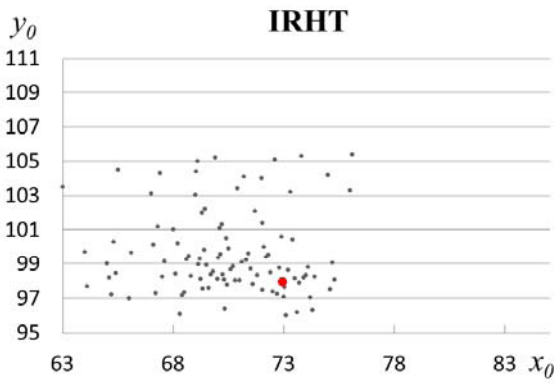
(a)

(b)

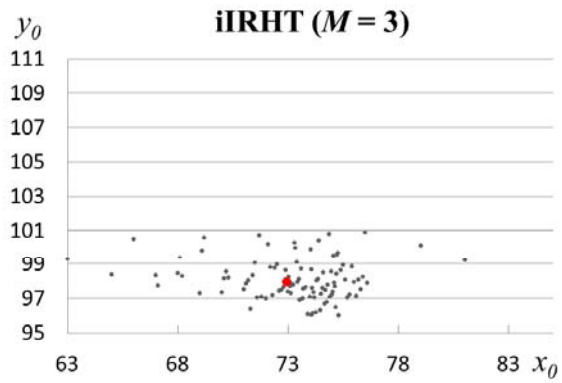
(c)


RHT
y_0
111
109
107
105
103
101
99
97
95
63
68
73
78
83
x_0

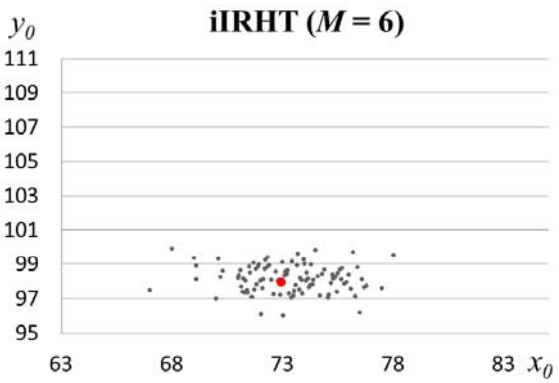
(d)


IRHT
y_0
111
109
107
105
103
101
99
97
95
63
68
73
78
83
x_0

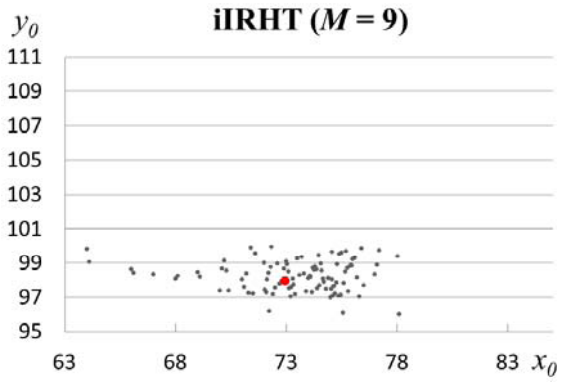
(e)


iIRHT ($M = 3$)
y_0
111
109
107
105
103
101
99
97
95
63
68
73
78
83
x_0

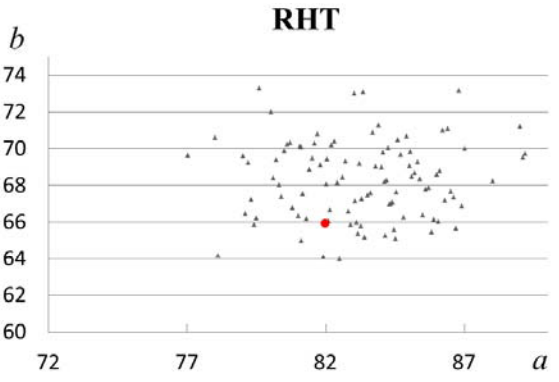
(f)


iIRHT ($M = 6$)
y_0
111
109
107
105
103
101
99
97
95
63
68
73
78
83
x_0

(g)


iIRHT ($M = 9$)
y_0
111
109
107
105
103
101
99
97
95
63
68
73
78
83
x_0

(h)


RHT
b
74
72
70
68
66
64
62
60
72
77
82
87
a

(i)

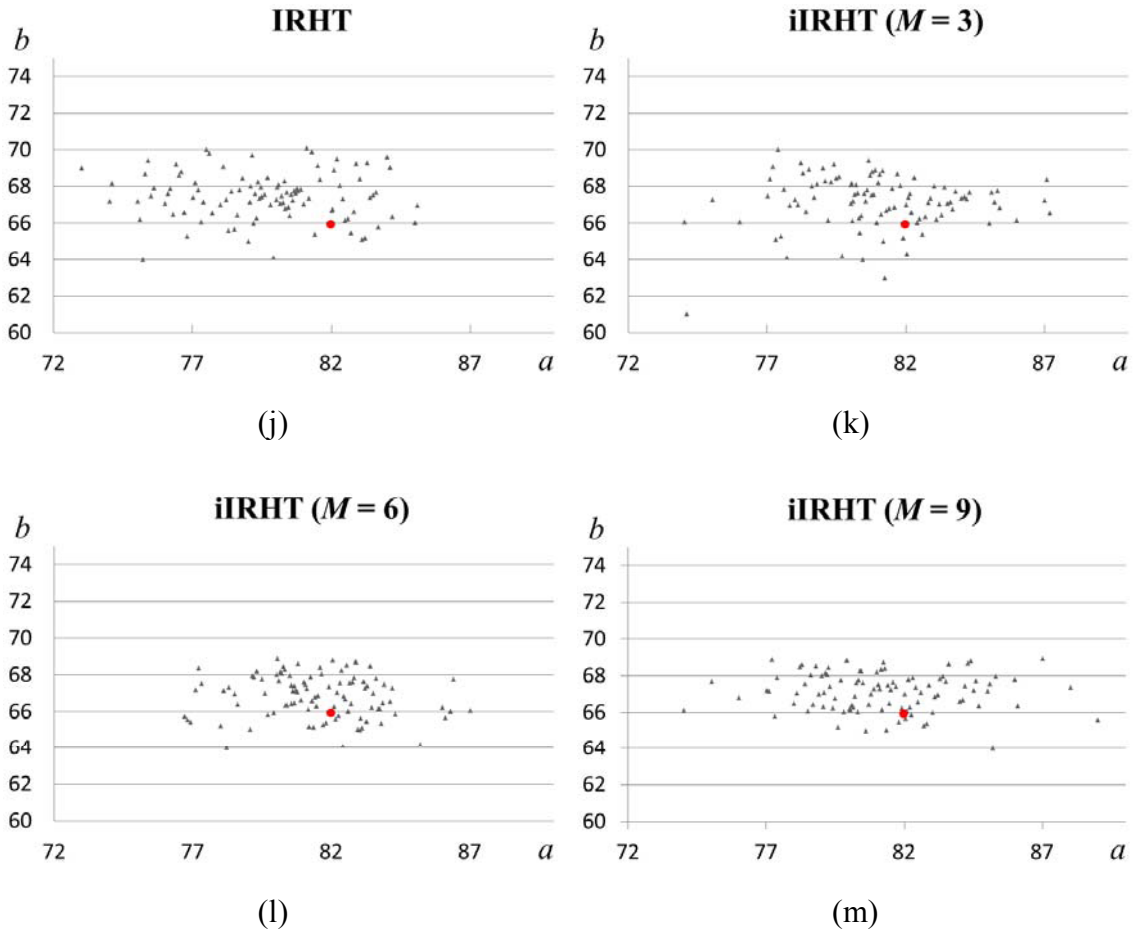


Figure 4.15 One US image and the detected results (a) a clinical fetal US image, (b) extracted skeletons, (c) reserved pixels by pre-processing from (b), (d)-(h) distributions of the centers of the detected ellipses by RHT, IRHT, iIRHT (*M* = *3*), iIRHT (*M* = *6*), and iIRHT (*M* = *9*), (i)-(m) distributions of the major and minor axes of the detected ellipses by RHT, IRHT, iIRHT (*M* = *3*), iIRHT (*M* = *6*), and iIRHT (*M* = *9*).

(1-2) A Test Database

In this sub-section, thirty-five clinical US images are utilized as a test database to compare the performance of three different methods, i.e., the RHT algorithm, the IRHT algorithm, and the proposed iIRHT algorithm (*M* = *6*), where six images are shown in Fig. 4.16 as examples. Since the parameters of the ellipse is five-dimension, after the ground truth of each image in the test database is obtained manually by the surgeons, the distance between the estimated parameters ($[x_0, y_0, a, b, \Phi]^t$) and the ground truth ($[x_g, y_g, a_g, b_g, \Phi_g]^t$) is calculated by

$$d^2 = (x_0 - x_g)^2 + (y_0 - y_g)^2 + (a - a_g)^2 + (b - b_g)^2 + (\Phi - \Phi_g)^2 \quad (4.18)$$

Then, the RMS value of a set of d's is computed by

$$RMS = \sqrt{\frac{1}{n}(d_1^2 + d_2^2 + \cdots + d_n^2)} \tag{4.19}$$

where, n is the test number of each image for each method, and herein n is selected as 50. Consequently, the mean RMSs and the mean execution time Ts of the algorithms are calculated based on the test database for each method, and listed in Table 4.5.

From Table 4.5, the proposed iIRHT method achieves the best accuracy, robustness and efficiency of the algorithm, compared with the RHT method and the original IRHT method. However, the mean RMS of the RHT method is extremely large, which proves it may fail when the image noise is very strong.

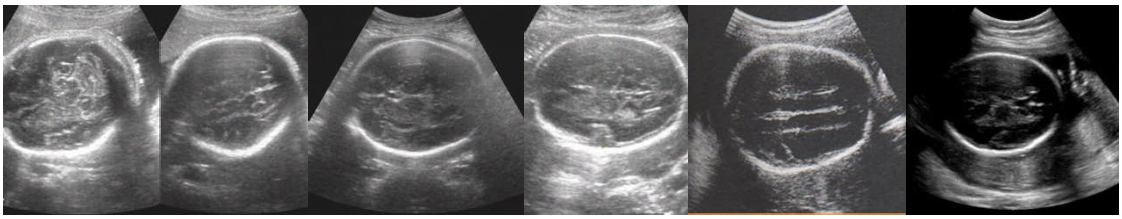


Figure 4.16 Six clinical US images.

Table 4.5 Comparison of three methods.

Methods	Mean RMS	Time (s)
RHT	98.84 ± 14.79	4.97 ± 3.36
IRHT	11.14 ± 2.46	3.45 ± 2.52
iIRHT	5.75 ± 0.68	1.24 ± 0.81

4.2.4 Conclusions

An improved iterative randomized Hough transform (IRHT) method has been proposed for automatic fetal head detection on ultrasound images. At first, the main skeletons of the object are extracted from images by the edge extraction. Then, the pre-processing is applied to remove redundant pixels and exclude edges that cannot generate one elliptic hypothesis with the selected edge by edge grouping. Meanwhile, the quasi random scheme, the least square approach, and the number of pixels on the ellipse are introduced to improve the proposed method. The experiments on clinical

fetal ultrasound images demonstrate that the proposed iIRHT method achieves more robust and accurate results, and has a better efficiency and performance than the RHT method and the original IRHT method.

4.3 A Novel 3D Ultrasound Calibration System

4.3.1 Introduction

In recent decades, much research has been undertaken to make ultrasound calibration more reliable, accurate, and simultaneously easier and quicker to perform. With the commercial availability of a new generation of 3D US machines, real-time 3D US calibration based on real-time 3D US volumes offers many new opportunities for computer-assisted diagnosis, therapy, and intervention. Usually, calibration methods are classified with respect to the target geometry (a phantom) they rely on. 3D US calibration methods using a 3D US probe have been reviewed specifically in [127, 128].

The simplest phantoms are probably single-point and multi-point targets. This can be formed by a pair of cross-wires or a stylus [127]. Poon and Rohling [127] compared three phantoms, namely an IXI-fiducial wire phantom, a cube phantom, and a stylus based on a 3D curved-array US probe. They obtained the root mean square (RMS) errors of the point accuracy measure of 2.15 mm (IXI-wire), 4.91 mm (cube), and 2.36 mm (stylus) through an optical tracking system. Lang et al. [129] examined two similar solutions based on 2D US images [130, 131] extended to 3D using EM tracking, and obtained an average distance error of 1.39 ± 0.54 mm. Nevertheless, most of them share the alignment and feature extraction problems.

Z-fiducial (N-fiducial) phantom addresses the alignment problem of point target methods. Bouchet et al. [132] designed a Z-phantom consisting of 39 nylon wires based on swept-volume US and an optical tracker. The maximum errors in the three dimensions of the volume are 0.4 mm in the elevation direction, 1.5 mm in the lateral direction, and 1.1 mm in the axial direction, leading to a total target location error of 1.9 mm. However, this technique required an average of 300 points for 3D US volumes, which made the calibration system inefficient.

Plane phantom methods are based on detection of the intersecting line of a planer surface with the 2D US beam. Michael et al. [133] optically tracked a single plane

membrane phantom, and achieved an RMS distance error of 1.15 mm. However, the major drawback of single-plane phantoms resides in their barely discriminative geometry. Meanwhile, the 12-step acquisition protocol is necessary for obtaining robust results, which leads to a lot of redundancy. Another drawback is that the overall calibration time is about 20 minutes, and the feature extraction problem results in a visual verification and eventually a correction.

Registration-based methods rely on surface or intensity based registration techniques and therefore can be performed on images of objects with arbitrary shape. Lange and Eulenstein [134] compared two image registration-based methods based on swept-volume 3D US, i.e., the tracked phantom (TP) approach and hand-eye (HE) calibration. The TP approach registered a US volume of a phantom with its associated CT volume, and the HE calibration registered two US volumes acquired from two different directions on the same phantom. They yielded an RMS location error of 3.3 mm with a magnetic tracking system. Huang et al. [135] described an image registration based 3D TEE-EM (transesophageal echocardiography) calibration system, and achieved an FRE (fiducial registration error) of 1.23 ± 0.26 mm and an average TRE (target registration error) of 2.37 ± 0.81 mm. Parthasarathy et al. [136] fused a pre-procedural MR roadmap with real-time 3D TEE for cardiac stem cell therapy. They validated the system on a moving heart phantom and produced a landmark registration accuracy of 2.8 ± 1.45 mm and an average registration accuracy of 2.2 ± 1.8 mm through EM tracking. Nevertheless, a robust registration procedure is often presumed for such methods, and sometimes a favorable matching process cannot be promised, which will cause a poor performance of the registration-based methods.

Some phantom-less methods are reported to solve 3D US calibration by locating the tip of a tracked stylus or needle precisely in 3D US volumes [137]. However, their validation on the calibration results was based on precision and not accuracy, and it was difficult to locate the needle tip accurately owing to 3D image resolution and artifacts.

In previous research, the high accuracy of point-based calibration methods has been demonstrated by several research groups [138-140]. Therefore, a novel point-based cones' phantom consisting of twelve resin cones is proposed for real-time 3D US calibration by EM tracking, which addresses the alignment and feature extraction problems, though one single cone phantom [141] has been used for free-hand 3D US calibration. In contrast to existing 3D US calibration systems, the feature extraction of

the proposed system is fully automated so that fiducial points can be segmented precisely in 3D US images, and the results prove that using only two US volumes can achieve high accuracy and efficient calibration.

4.3.2 A New Cones' Phantom

3D US calibration aims to determine the spatial transformation for mapping points from the 3D US space to the 3D tracking space. The transformation comprises six parameters: three translations in the direction of the *x*, *y*, and *z* axes and the three rotations, azimuth, elevation, and roll, about these axes.

Figure 4.17 shows the proposed calibration setup, in which twelve plastic cones, whose tips are utilized for the calibration, are placed on the bottom of a water tank. The 3D US probe with a 3D tracking sensor is fixed above the phantom by a plastic holder, so that the beam surface contacts the water. The transmitter of the 3D EM device is settled near the water tank. The *x*-*y*-*z* system is the coordinate system of 3D US images, the *x′*-*y′*-*z′* system is the coordinate system fixed to the 3D US probe, and the *x″*-*y″*-*z″* system is the world coordinate system of the 3D EM transmitter.

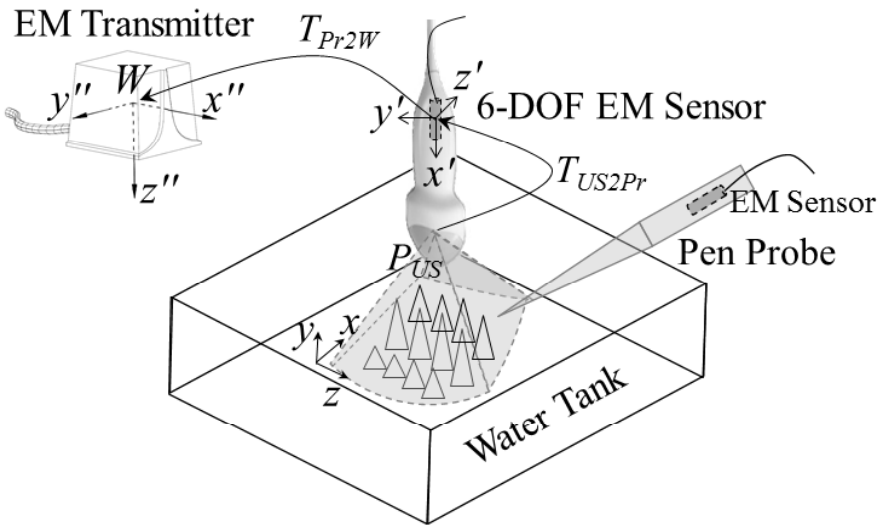
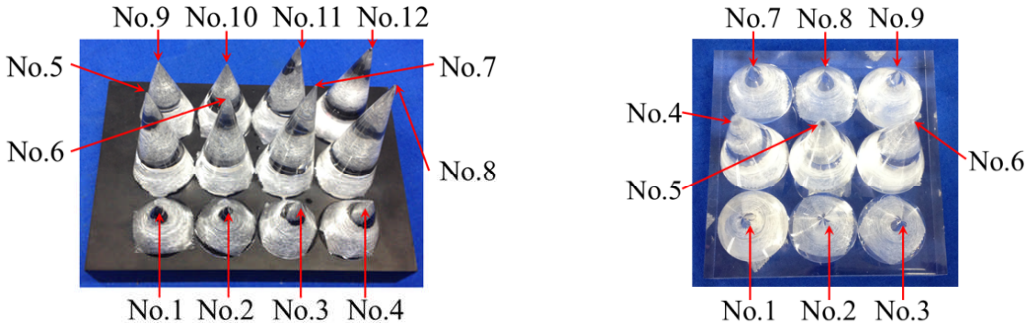


Figure 4.17 The experimental setup of the 3D US calibration system.

The phantom is composed of twelve transparent resin cones of six different types. The bottom diameters of all types are 30 mm, and the heights of the six types are 30 mm (C-I), 40 mm (C-II), 50 mm (C-III), 60 mm (C-IV), 70 mm (C-V), and 80 mm (C-VI),

respectively. Since the field of view of 3D US images is fan-shaped, its top is narrow and its bottom is wide. Accordingly, in the phantom, as shown in Fig. 4.18 (a), the two tallest cones are arranged in the middle, and other lower cones are arranged on both sides, i.e., the cones of No.5 and No.6 are C-V, No.7 and No.8 are C-VI, No.1 and No.2 are C-I, No.3 and No.4 are C-II, No.9 and No.10 are C-III, and No.11 and No.12 are C-IV. In addition, as shown in Fig. 4.18 (b), nine cones are arranged as a target model used for the validation, where the cones from No.4 to No.6 are C-IV (the taller ones in the middle), and the others are C-I (the lower ones on both sides).



(a) (b)

Figure 4.18 Experimental setup of the 3D US calibration system.

4.3.3 Calibration Matrix

The proposed 3D US calibration system addresses the feature extraction problem, i.e., the fiducial points (the tips of the cones are shown as white points in Fig. 4.19 (a)) are identified automatically and precisely in 3D US images by the following steps.

Step-1: The 2D slice of $y = 130$ (about the quarter position in the y direction) is extracted from 3D US images. It includes the twelve ellipses that correspond to the twelve cones in the phantom (Fig. 4.19 (b)). Note that in the 2D slice the cross section of each cone is observed as an ellipse due to US deformation.

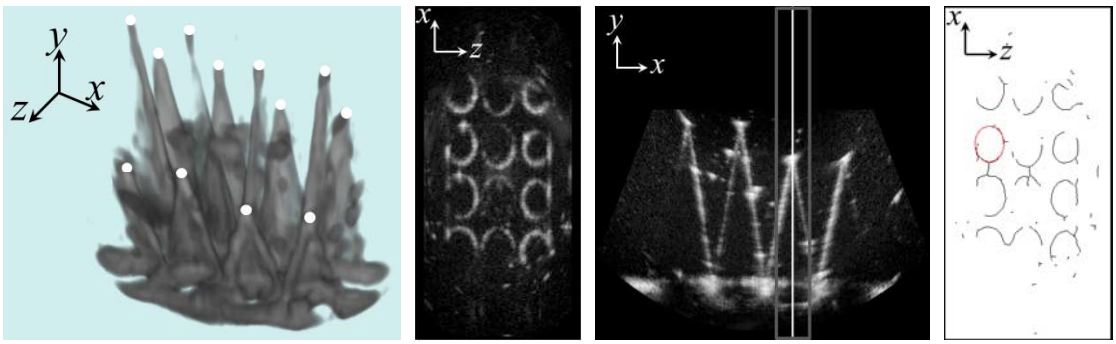
Step-2: Through the bilateral filter, the iKFCM segmentation [69], and distance transformation in the 2D slice, the skeletons of the twelve ellipses are extracted (Fig. 4.19 (d)).

Step-3: The twelve ellipses are detected by the improved IRHT method [118, 119] with the restrictions of $a \leq 43$, $b \leq 43$, and $e \leq 0.7$, where the semi-major axis a and semi-minor axis b of each cone should be less than $256/6 = 43$ (256 is the

maximum value of the *z*-axis), because a total of three cones are arranged in *z* direction, and *e* denotes the eccentricity of the detected ellipse. Since the US deformation is limited, the ellipses deformed from the circles of the cross sections of the cones should not be so flat that the eccentricity of the ellipses could be confined in the range of $e \leq 0.7$. Multiple-ellipse detection is accomplished by the method in [115], and then five elliptical standard parameters $[x_0, z_0, a, b, \Phi]$ ($a > b$) of each ellipse are estimated.

Step-4: Similarly, from the five 2D slices in *y* direction (i.e., *y*=130, 135, 140, 145, and 150), five sets of the twelve ellipses are extracted, and then the average values of the five parameters for each ellipse are calculated to suppress detection errors.

Step-5: To detect each cone's tip and identify which ellipse that detected tip belongs to, each ellipse's average center of $[\bar{x}_{ep}, \bar{z}_{ep}]$ and average semi-major axis $\bar{a}_{ep}$ (e.g., the red ellipse in Fig. 4.19 (d)) are computed, where, *ep* is the ID number of the twelve cones. In the 2D slice of $z = \bar{z}_{ep}$, a rectangular ROI (the gray rectangle in Fig. 4.19 (c)) is estimated, where the ROI's centerline is the line of $x = \bar{x}_{ep}$ (the white vertical line in Fig. 4.19 (c)), and length and width are 512 (the maximum value of the *y*-axis) and $\bar{a}_{ep} \times 2$, respectively. In the ROI area, the two lines of the edges of the cone are detected using Hough transform, so that the cone's tip is detected as the cross-point of the detected two lines. However, sometimes only one edge of the cone can be found (e.g., the rightmost cone in Fig. 4.19 (c)). In this case, the cross-point of the detected one edge and the line of $x = \bar{x}_{ep}$ is detected as the cone's tip. Similarly, the other cones' tips are detected. Then the 3D location of each tip in the phantom P_{US} is decided automatically and precisely from the 3D US volumes.



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

Figure 4.19 Automatic fiducial point detection in 3D US images. (a) 3D US volume, (b) 2D slice on x-z plane, (c) 2D slice on *x*-*y* plane, (d) skeletons of image (b).

On the other hand, the position P_{tip} of each tip in the phantom is manually measured by a pen probe, as shown in Fig. 4.17, where P_{tip} denotes the coordinate of each cone's tip in the 3D tracking space. The calibration matrix T_{US2Pr}, and the reference tracker's transformation matrix T_{Pr2W} have the following relationship:

$$P_{tip} = T_{pr2W} \cdot T_{US2pr} \cdot P_{US} \tag{4.20}$$

where $P_{US} = (x,\ y,\ z,\ 1)^T$, x and y are the column and row indices of the pixel on the extracted slice in x-y plane, and z is the index of the extracted slice along the z-axis. The scale factors are integrated with the calibration matrix T_{US2Pr}; therefore, the extra scale factor in the computation is not required. After being multiplied by the inverse of the reference transformation matrix T_{Pr2W} from left, Eq. (4.20) can be written as :

$$P_{tip/pr} = T_{pr2W}^{-1} \cdot P_{tip} = T_{US2pr} \cdot P_{US} \tag{4.21}$$

where P is a vector in the 3D US probe coordinate system. Considering n positions of the vertices in the phantom, the following equation is obtained from Eq. (4.21):

$$P_{tip/pr,1}, \cdots, P_{tip/pr,n} = T_{US2pr} \cdot \begin{pmatrix} x_1 \cdots x_n \\ y_1 \cdots y_n \\ z_1 \cdots z_n \\ 1 \cdots 1 \end{pmatrix} = T_{US2pr} \cdot T_{US} \tag{4.22}$$

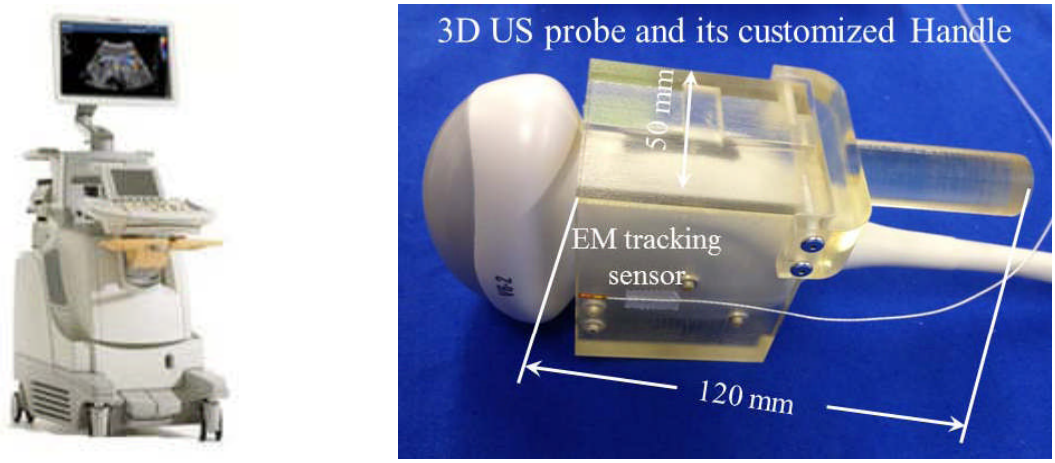
where $T_{US} = \begin{pmatrix} x_1 & \cdots & x_n \\ y_1 & \cdots & y_n \\ z_1 & \cdots & z_n \\ 1 & \cdots & 1 \end{pmatrix}$. Then, the calibration matrix is calculated by Horn's method [130, 142], as follows:

$$T_{US2pr} = (P_{tip/pr,1}, \cdots, P_{tip/pr,n}) \cdot T_{US}^T \cdot (T_{US} \cdot T_{US}^T)^{-1} \tag{4.23}$$

4.3.4 Results and Discussion

In practice, a Philips iU22 US system (Fig. 4.20 (a)) with a V6-2 3D curved-array US probe (Fig. 4.20 (b)) is employed to collect 3D US data, since curved arrays offer

significantly improved sizing than matrix arrays. The iU22 US system supports high resolution 2D/3D/4D imaging, and have a 6 to 2 MHz extended operating frequency range and a field of view of 66°. The Ascension 3D guidance trakSTAR (Fig. 4.21 (a)) with some Model 90 6-DOF tracking sensors (Fig. 4.21 (b)) is utilized for tracking, which can realize fast 3D tracking (240-420 updates per second). The diameter of the 3D tracking sensor is just 0.9 mm, so that it can be inserted into the surgical tool developed by Zhang et al. [26], whose inner diameter is only 1.0 mm. The static accuracy of the position for the tracking sensor is 1.4 mm RMS and the orientation is 0.5 RMS. The specifications of the Ascension 3D guidance trakSTAR is listed in Table 4.6. Moreover, all programs of each process in C++ run on a quad-core 2.60 GHz computer with 8GB memory.



(a) (b)

Figure 4.20 Ultrasound system.

(a) (b)

Figure 4.21 A 3D EM tracking device.

Table 4.6 Specifications of the Ascension 3D guidance trakSTAR.

Parameter	Detail	
Company	Ascension	
Type	Model 90 Sensor	
Size	**Sensor Housing:**	Outside Diameter: 0.9 mm (0.035 inch) Length: 8.8 mm (0.35 inch)
	Sensor Cable:	3.3 m (10.8 ft) distance
	Wide Range Transmitter:	12” (30.5cm) cube with 20’ (6.15m) cable
Degrees of freedom	Six (position and orientation)	
Translation range	±2.1m in any direction	
Angular range	All attitude: ±180° azimuth and roll, ±90° elevation	
Update rate	up to 600 updates/second	
Outputs	X, Y, Z positional coordinates, orientation angles, orientation matrix and quaternion	
Static accuracy	3.8mm RMS position	
	0.5 degree RMS orientation	
Interface	USB / RS232	
Power	100 to 240V, 50/60 Hz	
Operating temperature	59°F to 95°F (15°C to 40°C), 95% non-condensing humidity	

(1) Calibration Precision

Validation of the calibration is very necessary to evaluate the performance of the 3D US calibration in reconstructing the 3D plane in the tracking space. There are two common evaluation criteria for measuring calibration errors: (1) fiducial registration error (FRE) is the RMS distance between the located position of each fiducial as transformed from the US image space to the tracking space and the position of that corresponding fiducial located in the tracking space, which is used to evaluate how well the EM and US points fit together; (2) target registration error (TRE) is computed in the same manner as FRE, but the points for TRE are not used to estimate the calibration matrix, and thus it provides a better indication of the accuracy of the calibration.

Twenty-four fiducial points from 3D US images and tracking information were recorded, where 12 points are detected from one US volume by using the proposed phantom, and the other 12 points are detected from one more US volume by using the proposed phantom whose base is rotated by 180°. Additionally, 18 points from 3D US images are used for TRE measurement, where nine points are detected from one US volume by using the target model in Fig. 4.18 (b), and the other nine points are detected from one more US volume by using the same target model with some horizontal translation.

During the calibration, six points (the tips of the phantom from No.1 to No.6) are used initially, increasing by one point at a time up to 24 points (the point is added sequentially as the marked number increases from small to large, where the rotated phantom is marked in the same order) for FRE validation. For each set of calibration points, 18 points based on the target model are used for TRE validation. In total, the data of 10 groups (the data of the same experiment for ten times) are recorded, and each of which corresponds to 24 fiducial points and 18 target points. Additionally, the locations of the nine points phantom during this experiment are different, since the locations of the nine points phantom are considered as being changed horizontally and vertically so that the nine points phantom can cover the entire 3D US space. By doing this, location errors caused by image distortion on distal region of 3D US images can also be included into TRE.

The means and standard deviations of the 10 groups are calculated and illustrated in Fig. 4.22, with the FREs and TREs at each number of calibration points. As the number (n) of calibration points increases, it can be found that: (1) the means and standard deviations of FREs and TREs have a sudden decrease at $n = 9$ (the first black vertical line in Fig. 4.22); (2) those values start to converge at $n = 21$ (the second black vertical line in Fig. 4.22). As a result, the minimum number of the points used for the calibration system is 9. Moreover, more points (22 or more) do not cause much improvement for calibration results. Thus, 24 fiducial points from two US volumes are chosen to estimate the calibration matrix for locating the 3D fetal model with the oral cavity and airways, and the corresponding FRE of $1.49 \pm 0.44\,mm$ and TRE of $1.81 \pm 0.56\,mm$ are achieved in this case. Moreover, though feature extraction is computed in tens of seconds, the overall calibration time of about 6 minutes is mostly due to the manual measurement of the tip of each cone by a pen probe.

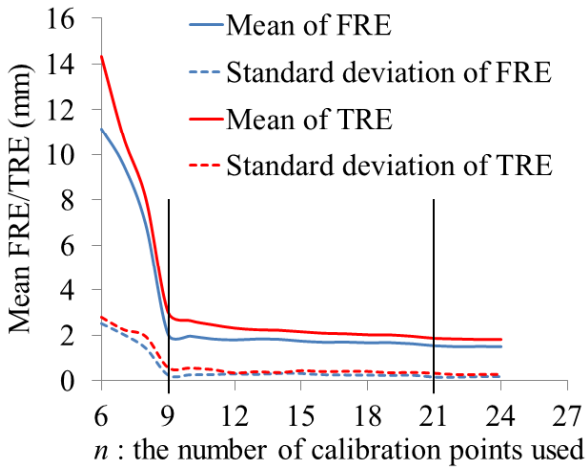
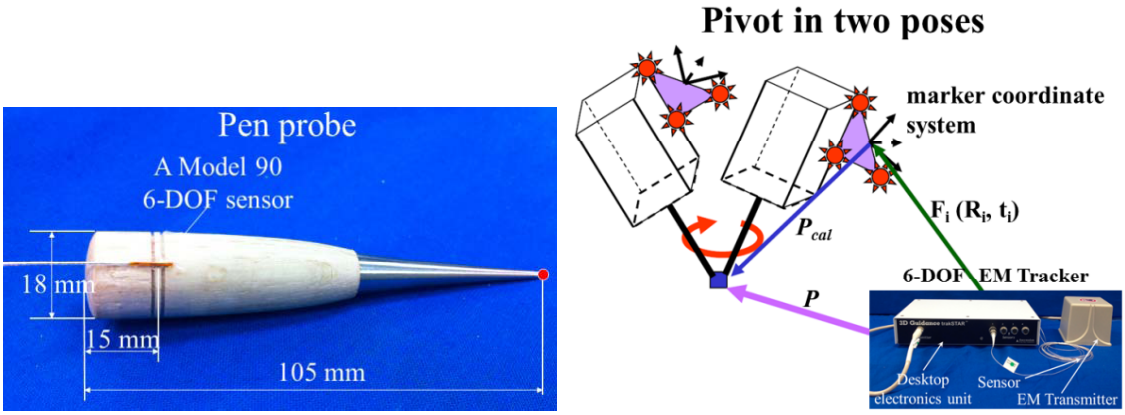


Figure 4.22 The validation results of the 3D US calibration system.



(a) (b)

Figure 4.23 A pen probe (a) and its transformation estimation (b).

Table 4.7 Comparison of the results of previous studies and the proposed approach.

Methods	**Results (RMS)**
Lange and Eulenstein	3.3 *mm*
Huang et al.	FRE: $1.23 \pm 0.26\ mm$ TRE: $2.37 \pm 0.81\ mm$
Parthasarathy et al.	$2.2 \pm 1.8\ mm$
Lang et al.	$1.39 \pm 0.54\ mm$
The proposed approach	**FRE: $1.49 \pm 0.44\ mm$ TRE: $1.81 \pm 0.56\ mm$**

Here, the pen probe is shown in Fig. 4.23 (a), which is attached by a Model 90 6-DOF tracking sensor, and is used to measure the positions of the objects in 3D tracking space (the coordinate system of EM transmitter). Unknown transformations of

the pen probe is estimated by pivot calibration, in which both the tip location relative to the tracked coordinate system and the fixed point's coordinates in 3D tracking space are computed by rotating the pen probe while its tip location remains fixed, as shown in Fig. 4.23 (b). Although two different poses are sufficient to calculate the unknown transformations of the pen probe, twenty different poses are used to take the average value of the transformations.

(2) Discussion

The results of the proposed calibration method based on 24 fiducial points are comparable to those obtained by previous calibration methods using EM tracking. The accuracy of the proposed approach is better than the majority of the methods from the comparison in Table 4.7. Although Lang et al. [129] obtained better accuracy than did the proposed approach, a set of 30-50 volumes is required for each experiment to ensure the best accuracy, which made the calibration process long and tiresome. In conclusion, the proposed system guarantees satisfactorily accuracy, and also only requires a simple experimental setup and two US volumes for 3D US calibration.

4.4 Conclusions

This chapter presents several basic algorithms on image segmentation , ellipse detection, and 3D US calibration, which significantly contribute to the final performance of the system. The improved Kernel-based fuzzy C-means algorithm heights the segmentation accuracy for 2D US image segmentation. The improved iterative randomized Hough transform method reduces the detection error and enhances the efficiency of the algorithm for detection of fetal head in 2D US images, fetal mouth in fetoscopic images, and position of each cone in the 3D US calibration system. The 3D Ultrasound Calibration System promises a high location accuracy and robustness of the system for locating the fetal facial surface extracted from 3D US data, the 3D fetal model with oral cavity and airways mapped with the extracted fetal facial surface, and the surgical tool detected from 3D US data.

Chapter 5 3D Fetal Face Detection

5.1 Introduction

In the last few years, ultrasound (US) images have been widely used for visualizing and diagnosing internal organs and fetuses due to their safety, cost efficiency and impact-free characteristic compared to other modalities, e.g. MRI and CT. Real-time 2D and 3D US images have become important aids for guiding a minimally invasive surgery [33], e.g., as described in Chapter 1 , fetoscopic tracheal occlusion (FETO) [9] surgery, which is operated to treat a severe birth defect of the fetuses called congenital diaphragmatic hernia (CDH). In order to navigate such a surgery, the detection of 3D fetal face and location of key facial features from 3D US data, e.g. nose, eyes, lips and mouth, could be very helpful for reducing the surgical risk.

On the other hand, the ordinary 3D sensor-based 3D face recognition technology has been developed for many critical applications over the past decades, such as human computer interface, personal authentication, surveillance systems, etc. Some earlier works on curvature analysis [78-81] are proposed for face recognition based on distinct facial curvature features. 3D model-based methods [82, 83] are often employed for face recognition by fitting a priori 3D face model to a given face. In addition, some methods represent the 3D data in a different domain for face recognition, e.g. 3D PCA [84], shape index [85], point signature [86], and spine image [87], etc. Some other methods, e.g. face profile extraction [88], and facial curves [89], are also reported recently. Several surveys of 3D face recognition methods can be found in [143-145].

However, face recognition technologies are rarely exploited in medicine, especially for 3D US images. To our best knowledge, there is only one work to detect 3D fetal face from US volumes for the optimal view [90]. They proposed a learning-based approach by combining 3D and 2D information to detect fetal face from 3D US volumes in one second on a dual-core 2.0 GHz computer, but their method required a priori 3D face mesh model for face detection and a high calculation complexity on the registration between the 3D model and the object. Additionally, curvature-based surface descriptor is one of the most successful 3D surface representations for 3D face recognition. As a result, a curvature feature-based approach is proposed to automatically detect 3D fetal

face and accurately locate key facial features for guiding the FETO surgery in a high efficiency [146]. After that, the detected position of the upper lip in 3D US space is transformed into 3D tracking space by a 3D US calibration system, and then the physical distance between the detected upper lip and the tip of the surgical tool will be estimated to provide an alarm for surgeons in the future.

This work is different from the previous ones in several aspects. First, the proposed method does not require any priori knowledge or training data. Second, the scaling between a 3D face model and a given face is not necessary to be computed. Third, a new boosting traversal scheme based on the spatial relations between each facial feature is presented for 3D face detection and facial feature location. Fourth, this system has a high efficiency and takes about 625 ms for locating all key facial features on a quad-core 2.60 GHz computer.

5.1.1 Background

As described in Section 1.1.3, in an FETO surgery, a fetoscope is inserted through the abdominal wall into the uterus, and a detachable balloon is placed in fetal trachea for the tracheal occlusion by a flexible and slender surgical tool developed in [25], as shown in Fig. 5.1, to stimulate lung growth via increasing the pressure of fetal chest cavity. During the operation, the pose of a 3D US probe is easily controlled within 15° from the frontal view so as to allow an optimal view of the fetal face.

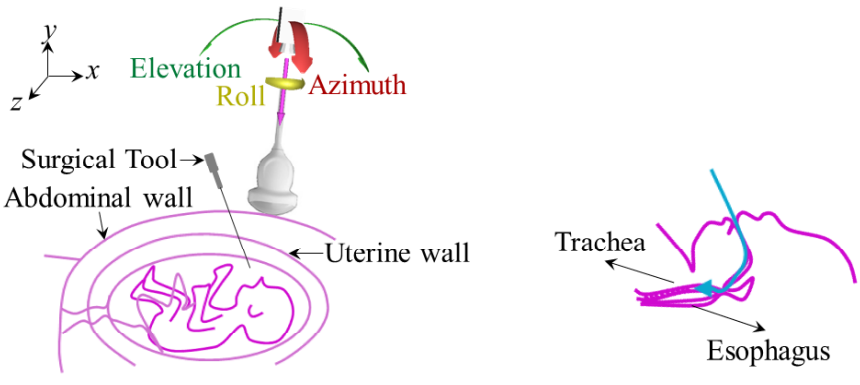


Figure 5.1 The FETO surgery.

Besides, the surgical tool is automatically operated by a robot control system during operations. In order to promise the surgical security, the movement accuracy of

the control system should be improved further, and the moving speed of the surgical tool should be reduced if the distance between the fetal mouth and the tip of the surgical tool is less than 1.5-2.0 cm. For such a purpose, automatic fetal face detection and accurate and precise location of the key facial features are very significant for the navigation system of the FETO surgery.

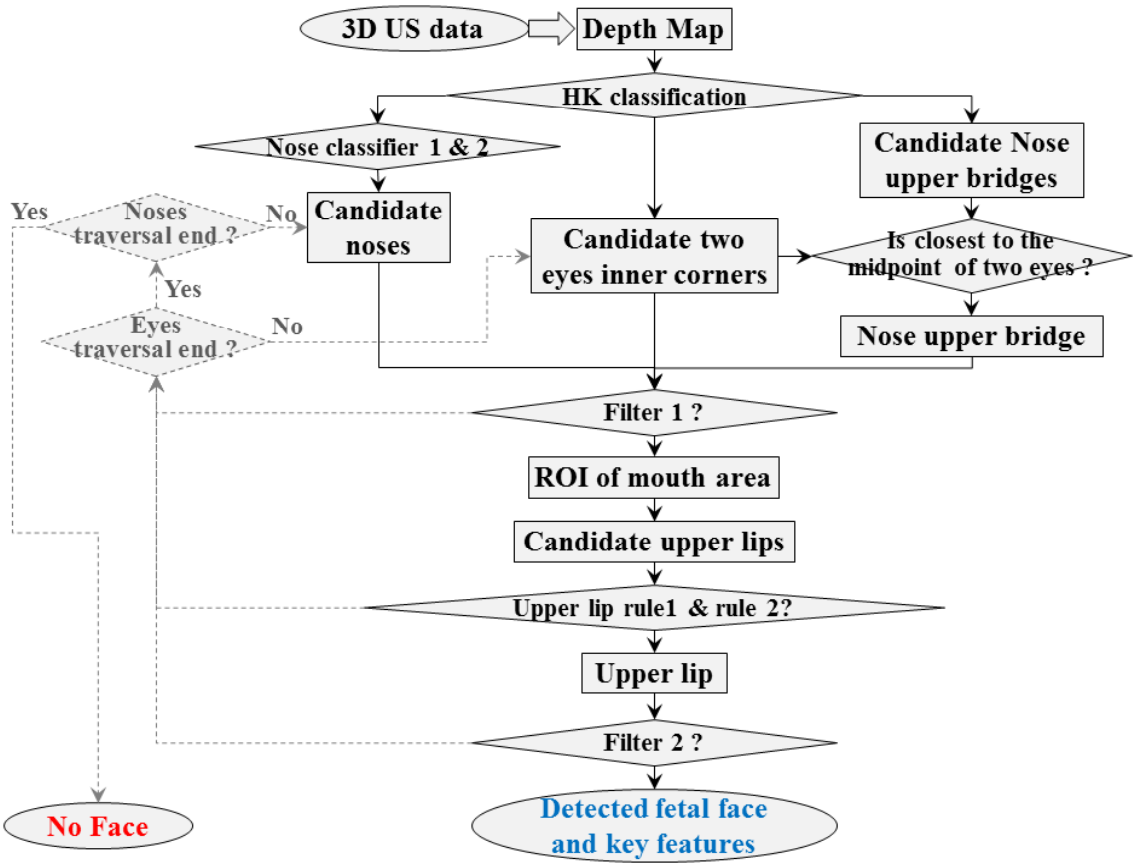


Figure 5.2 The flow diagram of the proposed approach.

5.2 Method

5.2.1 Process of the Proposed Method

In this chapter, the key fetal facial features are identified, including the nose tip, two eyes inner corners, nose upper bridge and upper lip, by the mean and Gaussian curvatures in the 3D fetal face extracted from 3D US volumes. Figure 5.2 shows the major steps of the proposed approach based on a boosting traversal scheme.

At first, the depth map of the fetal face is obtained from the 3D US volumes.

Second, the candidate regions of the nose tip, two eyes inner corners and nose upper bridge are segmented by the HK classification with different parameters. Third, the positional relations among four features are verified by the ***Filter 1***. If they meet the conditions of the ***Filter 1***, the region of interest (ROI) of the mouth area is estimated to identify the position of the upper lip; otherwise all possible combinations of four features among all candidate regions are traversed and validated by the ***Filter 1*** until an appropriate combination is found. Fourth, the candidate regions of the upper lip are separated by the HK classification in the ROI of mouth area, and one of them is extracted based on the *Upper lip rule 1* and *2*. Finally, the positional relations among all detected five features are checked by the ***Filter 2***. If the ***Filter 2*** is satisfied, then a fetal face with the key features is successfully detected; otherwise the traversal process on the other combinations of each feature continues until a conclusion is obtained.

5.2.2 Acquisition of Depth Image for Fetal Face

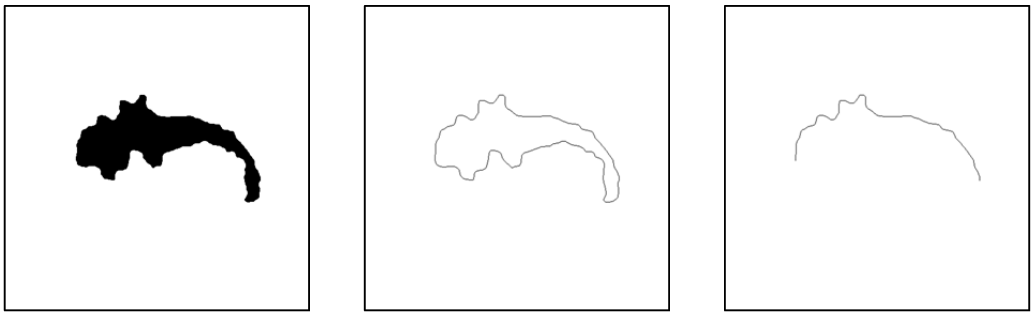
In experiments, a physical fetal phantom made of silicone resin is scanned in a water tank, by a Philips iU22 US system and a V6-2 3D US probe to obtain 3D US images with the size of $512(x) \times 510(y) \times 256(z)$. Herein, a fetal phantom at the gestational age (GA) of 32 weeks is taken as an example, and its 3D US data is shown in Fig. 5.3 (a). As we all know, 3D US images may have low resolution, and the imaging quality is easily affected by phantom material and air bubbles in the water. Therefore, a method is proposed to extract 3D fetal facial surface by scanning the slices along *z* axis. The procedure of the method is described in the following steps, where the slice of *z =128* is shown as an example in Fig. 5.3 (b).

Step-1 The boundaries in the 2D slice of *z =128* are detected by canny edge detection in Fig. 5.3 (c).

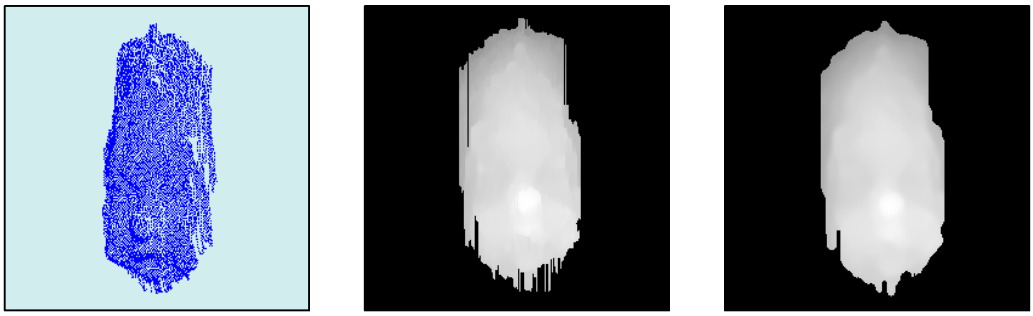
Step-2 The fetal head is detected by fitting an ellipse to the detected boundaries in Fig. 5.3 (c) via the improved iterative randomized Hough transform (IRHT) algorithm proposed in [118, 119] and Section 4.2 , so that five elliptical standard parameters $[x_0, y_0, a, b, \emptyset](a > b)$ are obtained. Next, a foursquare ROI with the center of (x_0, y_0) and the side length of $a \times 1.1 \times 2$ is selected;

Step-3 The largest connected region is thresholded in the ROI area and filtered by binary morphologic closing and opening operations and a hole filling filter, as shown in Fig. 5.3 (d);

(a) 3D US data (b) 2D slice (c) edges & ellipse



(d) the largest region (e) contour (f) upper edges



(g) 3D point cloud (h) depth image (i) smoothed result

Figure 5.3 Acquisition of depth image for fetal face.

Step-4 The contour of the largest connected region is detected by a binary contour filter in Fig. 5.3 (e), and the upper edges above the midline of the contour are accepted in Fig. 5.3 (f);

Step-5 Repeat the steps from (3) to (4) on the slices from $z = 50$ to $z = 200$ (because the fetal head is placed at the center in 3D US space), then a 3D point cloud of the fetal facial surface is extracted in Fig. 5.3 (g).

Step-6 Finally, the depth image of the fetal facial surface is mapped to 2D space, as shown in Fig. 5.3 (h).

This method does not consider the impact of obstructions in front of the face, e.g. fetal hands, because they will be moved away from the front of the face during the operation. In addition, the pose of the fetal face from the frontal view is controlled in the range of [0°, 15°] in operations, which means the elevation view angle, the azimuth view angle, and the roll view angle around nose tip (focal point) are all controlled in this range, as shown in Fig. 5.1 (a). If the pose of the fetal face is non-zero, the depth errors will be derived from the extracted 3D point cloud of the facial surface. However, the experimental results in Section 5.3 indicate that the depth errors are sufficiently small and do not affect the detection of 3D fetal face and key facial features for the proposed approach.

5.2.3 HK Classification

The HK classification method [147] is performed to obtain a point classification in a depth image according to the local shape around each point, where the mean (H) and Gaussian (K) curvatures respectively describe the property and bending degree of the local shape. The different local surface shapes in a 3D facial mesh can be determined by the signs of the mean and Gaussian curvatures as shown in Table 5.1.

Table 5.1 HK classification.

	H < 0	H = 0	H > 0
K < 0	hyperbolic concave	hyperbolic symmetric	hyperbolic convex
K = 0	cylindrical concave	planar	cylindrical convex
K > 0	elliptical concave	impossible	elliptical convex

To calculate the curvatures of 3D faces, let $S \subset R^3$ be a surface defined as follows:

$$S = \{(x, y, z) \mid z = F(x, y) = ax^2 + by^2 + cxy + dx + ey + f\} \quad (5.1)$$

For each point $(x, y, F(x, y)) \in S$, the mean (H) and Gaussian (K) curvatures [148] are computed by

$$H(x, y) = \frac{(1 + F_x^2)F_{yy} - 2F_xF_yF_{xy} + (1 + F_y^2)F_{xx}}{2(1 + F_x^2 + F_y^2)^{\frac{3}{2}}} \quad (5.2)$$

$$K(x,y)=\frac{F_{xx}F_{yy}-F_{xy}^2}{2(1+F_x^2+F_y^2)^{\frac{3}{2}}} \tag{5.3}$$

where F_x, F_y, F_{xy}, F_{xx}, F_{yy} are the 1st and 2nd derivatives of F in (x,y). Since there are some artifacts, noise and gaps in the depth image extracted from the 3D US volumes, a median filter and a bilateral filter are executed to remove sharp spikes and noise, and smooth the depth image, as shown in Fig. 5.3 (i).

5.2.4 Nose, Eyes and Nose Upper Bridge

As introduced in [80, 81], the shapes of the nose tip, nose upper bridge and two eyes inner corners are elliptical convex, hyperbolic convex, and elliptical concave, respectively. In the beginning, the depth image is thresholded to isolate regions of high curvatures, and the points with low curvatures are rejected by $|H(x,y)| \geq T_h$ and $|K(x,y)| \geq T_k$, where T_h and T_k are predefined and tested for each feature by experiments. Then the nose tip, two eyes inner corners and nose upper bridge are detected in the following sequence.

(1) Nose Tip

As the first step of the boosting traversal scheme, nose tip detection is extremely crucial to improve the efficiency and robustness of the proposed algorithm. There have been some works about the location of the nose tip, e.g. the highest point in the depth image [88], curvatures [149], and local features [150], etc. However, no method can promise a recognition rate of 100% for the location of the nose tip due to the noise and rotation of the subject. Here, a boosted classification scheme is proposed to identify the candidate regions of the nose tip based on the local curvature features.

Nose classifier 1: The candidate regions of the nose tip are distinguished by the HK classification ($H>0, K>0$) and the threshold ($|H| \geq 0.04, |K| \geq 0.0005$).

From the results of *Nose classifier 1* in Fig. 5.6 (a), morphological opening and closing operations are performed to exclude some small candidate regions (Fig. 5.6 (b)). In order to depict the cap-like shape of the nose tip, the average value and standard deviation of the Gaussian curvatures for each candidate region are calculated.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} K_i \tag{5.4}$$

$$\mu_j = \frac{1}{N_j}\sum_{i=1}^{N_j} K_{ij} \tag{5.5}$$

$$\sigma_j = \sqrt{\frac{1}{N_j}(\sum_{i=1}^{N_j} K_{ij}^2) - \mu_j^2} \tag{5.6}$$

where μ and N are respectively the average value of the Gaussian curvatures and the number of the points for all candidate regions, μ_j, σ_j and N_j are respectively the average value and standard deviation of the Gaussian curvatures and the number of the points in the j-th candidate region, K_i is the Gaussian curvature of point i, and K_{ij} is the Gaussian curvature of point i in region j. Ordinarily, the nose tip is the most distinct protuberance on the facial surface, i.e. the region of the nose tip has the largest average value of Gaussian curvatures. However, owing to the unpredictable sharp spike noise in the depth image, the average value and standard deviation of the Gaussian curvatures in the nose tip region are relatively larger than those of any other facial data.

Nose classifier 2: The candidate regions satisfying $\mu_j \geq \mu$ from the results of *Nose classifier 1* are remained, and sorted in descending order by the standard deviation σ_i.

Through the boosted classification scheme based on *Nose classifier 1* and *2*, a few candidate regions are chosen, and the points with the highest Gaussian curvature in each candidate region are selected as the candidate positions of the nose tip, e.g. the black points in Fig. 5.6 (c). Then the candidate points in the sorted order are traversed one by one with other features to check their spatial relations of the ***Filter 1*** and ***Filter 2***.

(2) Two Eyes Inner Corners

The regions of two eyes inner corners have the largest average value of the Gaussian curvatures in a clean depth image for elliptical concave areas, but this criterion is not suitable for a fetal facial surface since the fetal facial features except nose tip are not obvious compared with those features of the adult. Accordingly, the candidate regions of the two eyes inner corners are segmented by the HK classification ($H < 0, K > 0$), as shown in Fig. 5.6 (d). Some small candidate regions are discarded in Fig.

5.6 (e) by morphological opening and closing operations. Subsequently they are sorted in descending order by the standard deviation of the Gaussian curvatures of each candidate region. Then the points with the highest Gaussian curvature in each candidate region are selected as the candidate positions of two eyes inner corners, e.g. the black points in Fig. 5.6 (f). After that, all the sorted candidate positions are traversed in pairs as the order in Fig. 5.4 to combine with the detected nose tip for the examination of the ***Filter 1***, where n is the sum of the eyes' candidate positions.

(3) Nose Upper Bridge

The candidate regions of the nose upper bridge are classified by the HK classification ($H > 0, K < 0$) and the threshold of the depth map ($|H| \geq 0.005, |K| \geq 0.00005$), e.g. the grey regions in Fig. 5.6 (g). Then the candidate region closest to the midpoint of the two eyes inner corners is selected as the region of the nose upper bridge, whose centroid is calculated as the position of the nose upper bridge, e.g. the black point in Fig. 5.6 (g).

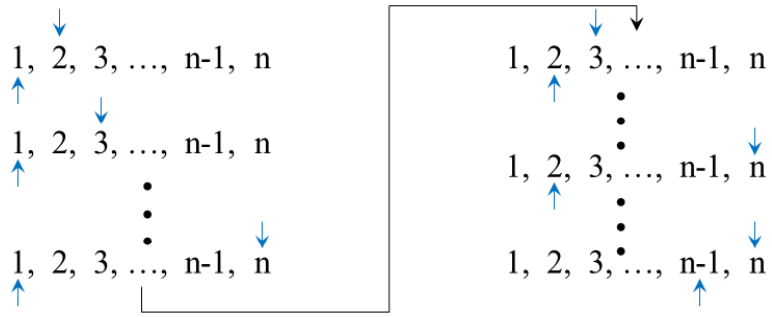


Figure 5.4 The selection order of the two eyes inner corners.

5.2.5 Filter 1

For the key facial features, there is a positional relationship among them, as shown in Fig. 5.5 (a), where n_1 is the position of the nose tip (green point), n_2 is the position of the nose upper bridge (light green point), e_1 and e_2 are the positions of the two eyes inner corners (two blue points), e_m is the midpoint between e_1 and e_2 (light blue point), c is the crosspoint between the line from n_1 to n_2 and the line from e_1 and e_2 (purple point), α_1 is the angle between the vector from e_1 to n_1 and the vector from e_1 to e_2, α_2 is the angle between the vector from e_2 to n_1 and

the vector from e_2 to e_1, β is the angle between the vector from n_2 to n_1 and the vector from e_2 to e_1, γ_1 is the angle between the vector from e_1 to n_2 and the vector from e_1 to e_2, and γ_2 is the angle between the vector from e_2 to n_2 and the vector from e_2 to e_1. Ideally, the above mentioned parameters should satisfy the following conditions:

① $\alpha_1 = \alpha_2 < 90^\circ$;

② $\beta \cong 90^\circ$;

③ $\gamma_1 < 90^\circ$, and $\gamma_2 < 90^\circ$ are restricted to guarantee the position of n_2 be between the two eyes inner corners;

④ The point *c*, point e_m and point n_2 are very close.

Actually, the corresponding parameters of each restriction are predefined and tested in experiments. Thus, the ***Filter 1*** is defined as follows:

① $\alpha_1 < 90^\circ$, $\alpha_2 < 90^\circ$, and $|\alpha_1 - \alpha_2| \leq 25^\circ$;

② $65^\circ < \beta < 115^\circ$;

③ $\gamma_1 < 90^\circ$, and $\gamma_2 < 90^\circ$;

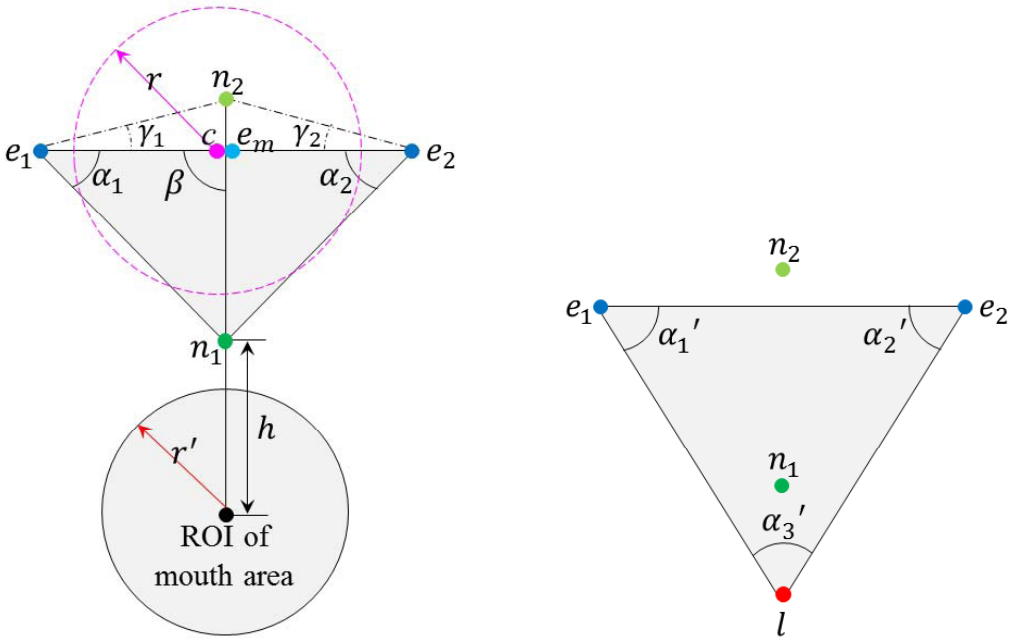
④ $d_{e_m \to c} < r$, $d_{n_2 \to c} < r$, and $d_{n_2 \to e_m} < r$, where $r = 0.75 \times min\{d_{n_2 \to n_1}, d_{e_2 \to e_1}\}$.

where $d_{e_m \to c}$ denotes the pixel distance between e_m and *c*, and other *d*-s have the same meanings. If the candidate positions of the nose tip, nose upper bridge and two eyes inner corners meet each condition of the ***Filter 1***, then a circle ROI of the mouth area is estimated to identify the position of the upper lip, as shown in Fig. 5.5 (a) and Fig. 5.6 (h), where the ROI's center (black point in Fig. 5.5 (a)) is away from n_1 by $h = 1.5 \times r'$, on the line from n_2 to n_1, and the ROI's radius $r' = max\{d_{n_2 \to n_1}, d_{e_m \to n_1}\}$, otherwise other combinations of candidate nose and eyes are traversed.

5.2.6 Upper lip

Though seven key features of a human face is possible to be detected by the local shape index [151], it is still difficult to detect some key features for a fetal face, such as

two eyes outer corners, two eyes, and two mouth corners, even if the local shape index is tried. Thus, it is forced to detect the most remarkable feature - the upper lip in the ROI of the moth area, whose shape is elliptical convex.



(a) (b)

Figure 5.5 The positional relationship among the detected key facial features. (a) the Filter 1, (b) the Filter 2.

Candidate regions of the upper lip are identified in the ROI of the mouth area by the HK classification ($H > 0, K > 0$), as shown in Fig. 5.6 (i). Small candidate regions are removed as shown in Fig. 5.6 (j), by morphological opening and closing operations. The centroids of the candidate regions are selected as the candidate positions of the upper lip, e.g. the black point in Fig. 5.6 (k). After then, the point satisfying the following rules is determined as the position of the upper lip.

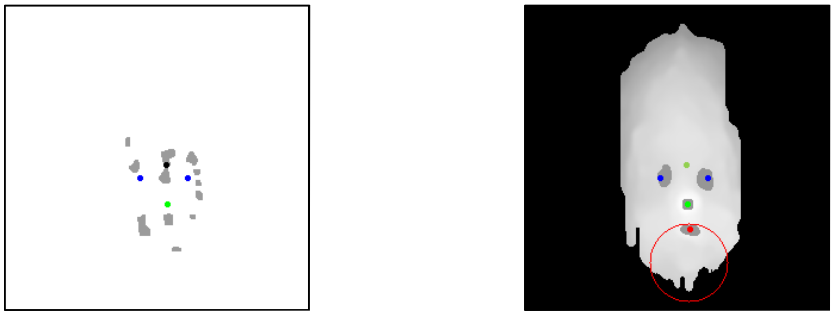
Upper lip rule 1: The angle between the vector from n_1 to the position of the upper lip l and the vector from n_2 to l is less than 10°, because l, n_1 and n_2 should be on the same line, as shown in Fig. 5.5 (b).

Upper lip rule 2: Among the remained points (as a result of Upper lip rule 1), the point closest to n_1 is determined as the position of the upper lip.

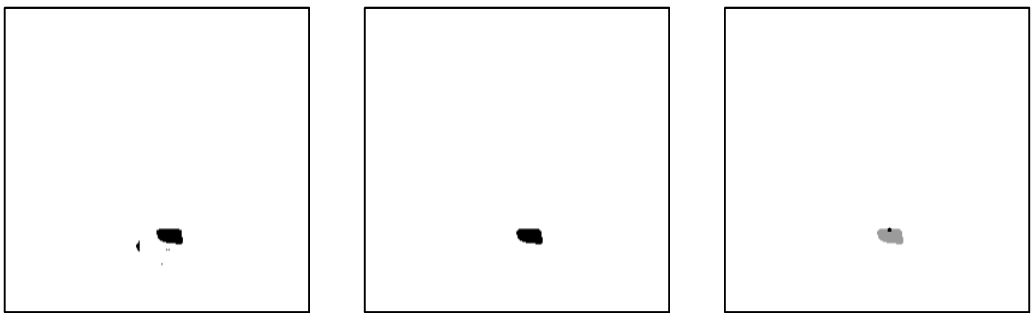
If there is no point that satisfies the *Upper lip rule 1* and *2*, then other combinations of candidate nose and eyes are traversed.

(a) nose candidates (b) morphological operations (c) nose positions

(d) eyes' candidates (e) morphological operations (f) eyes' positions



(g) candidates of nose upper bridge (h) detected feature points



(i) upper lip candidates (j) morphological operations (k) upper lip positions

Figure 5.6 The process of the fetal face and key features detection.

5.2.7 Filter 2

If the position of the upper lip l is detected by the Filter 1, then the spatial relationship among all key features in Fig. 5.5 (b) is examined by the ***Filter 2***. Thus, ***Filter 2*** is defined as follows.

① The triangle $\Delta e_1\, e_2\, l$ is an acute triangle, i.e. $0^\circ < \alpha_1' < 90^\circ$, $0^\circ < \alpha_2' < 90^\circ$, and $0^\circ < \alpha_3' < 90^\circ$;

② n_1 is inside the triangle $\Delta e_1\, e_2\, l$.

Qualified by both the ***Filter 1*** and ***Filter 2***, then the fetal face and the key facial features are judged to be successfully detected from the depth image; otherwise, other combinations of candidate nose and eyes are traversed until a conclusion is obtained.

5.3 Experiments

In experiments, 72 images of 6 subjects are collected to test the efficiency and robustness of the proposed algorithm. All programs run on a quad-core 2.60 GHz computer with 8GB memory. Since the FETO surgery is operated at about 26-29 weeks gestational age [23], 6 fetal phantoms with different sizes made by Kyoto Kagaku Co., Ltd are utilized as shown in Fig. 5.7, where the six phantoms are respectively at the gestational age of 24 weeks, 26 weeks, 28 weeks, 30 weeks, 32 weeks, and 34 weeks. Naturally, the larger the gestational age of the fetuses, the more distinct the fetal facial features. For each phantom, 12 typical poses are captured by rotating the 3D US probe around the nose tip, i.e. the frontal view, $\pm 10^\circ$ and $\pm 15^\circ$ of azimuth angles, $\pm 10^\circ$ and $\pm 15^\circ$ of elevation angles, and 30°, 45° and 60° of roll angles from the frontal view, as shown in Fig. 5.1 (a). Among such a test database, the ground truth is decided manually in 3D US images, as well as all key feature points are detected by the proposed algorithm for each case, and thus a detection success rate of 100% is achieved. Moreover, a detection success rate of 100% is obtained for any roll angle, because the roll rotation does not cause the errors of the extracted depth image. Figure 5.8 portrays some examples of several poses of the fetal phantoms with the detected key facial features. In addition, the proposed system has a high efficiency and can detect all key facial features in about 625 ms (the average execution time for all test data, as shown in Table 5.2). Compared with the efficiency (one second) of the fetal face detection

method in [90], the performance of the proposed system is elevated by 37.5% without any algorithm optimization; therefore, it is possible to speed up the system 10-20 times by utilizing GPU technology or parallel processing.

Figure 5.7 Six fetal phantoms.

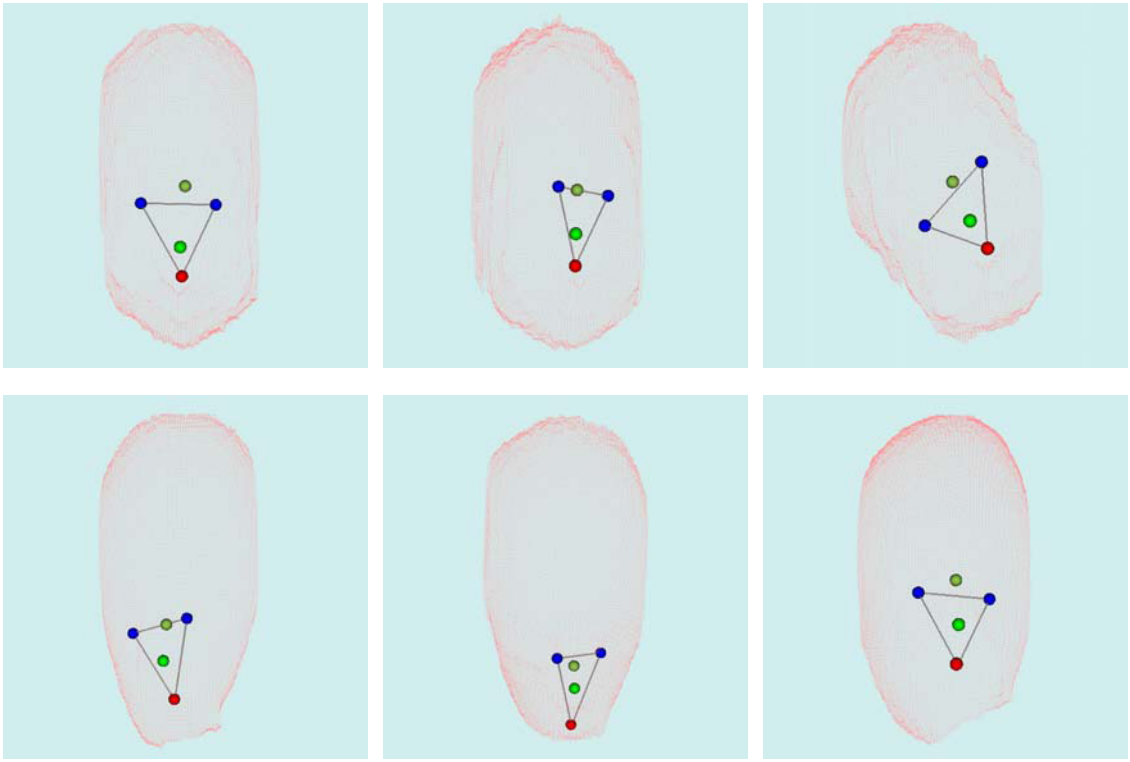


Figure 5.8 The results of the detected fetal face and key features.

On the contrary, there are some errors on the key feature location. Due to the significance of the upper lip location for the FETO surgery, the distance error of 9.13 ± 2.61 voxels for all tested images is calculated by measuring the detected upper lip and the marked upper lip manually, i.e. the root mean square (RMS) distance error of $3.18 \pm 0.91\ mm$ by a transformation from 3D US space to 3D tracking space according to the 3D US calibration system. Subsequently, the distance between the

detected upper lip and the tip of the surgical tool will be measured to provide an alarm for surgeons when it is less than $1.5 - 2.0\ cm$. Therefore, the distance error of the detected upper lip can be tolerated by the FETO surgery.

Table 5.2 The comparison results.

	HK Classification	**SC Segmentation**	**Principal Curvature**
Distance error	$3.18 \pm 0.91\ mm$	$3.26 \pm 0.85\ mm$	$3.84 \pm 1.37\ mm$
Efficiency	$625 \pm 50\ ms$	$642 \pm 46\ ms$	$1109 \pm 132\ ms$

In Table 5.2, the results about the distance errors and the efficiency of the method are compared by three different curvatures, i.e. the mean (H) and Gaussian (K) curvatures, the shape index (S) and the curvedness value (C) [85, 152], and the maximum and minimum principal curvatures [78, 79]. It can be learned that the performance of the HK classification and the SC segmentation on the tested images is basically the same, but the principal curvature does not perform as good as the previous two methods owing to the lack of the curved-ness representation. So the principal curvature-based method cannot identify the key facial features with a small distance error and a high efficiency. Accordingly, the HK classification and the SC segmentation both are effective and robust for detecting 3D fetal face and locating the key facial features.

5.4 Conclusions

In this chapter, a novel boosting traversal scheme based on the mean and Gaussian curvatures of the surface has proposed to detect 3D fetal face and locate the key facial features from 3D US data for navigating an FETO surgery. The experimental results indicate the proposed approach is efficient and robust for 3D fetal face detection and feature location, i.e. all key feature points are detected for each case, and thus a detection success rate of 100% is achieved by testing 72 images of 6 subjects, and the location error $3.18 \pm 0.91\ mm$ of the detected upper lip can be tolerated by the FETO surgery, as well as the system can complete the entire detection process in about 625 ms on a quad-core 2.60 GHz computer. In the future, the proposed approach will be improved to reduce the upper lip location error by optimizing the parameters of each filter and classifier according to more clinical fetal ultrasound data. Meanwhile, the

depth image extraction method will further be developed to acquire more accurate depth information for 3D fetal face detection and facial feature location.

In this thesis, this algorithm is applied for detection of key features on 3D fetal face, where the detected key features are utilized for coarse registration between the 3D fetal model and the extracted 3D fetal facial surface (Chapter 7), and the detected lip (one key feature) will be utilized for estimating the distance between the tip of the surgical tool and the detected lip in the future.

Chapter 6 Surgical Tool Location and Fetal Mouth Tracking

6.1 Surgical Tool Location

This section proposes an approach that accurately localizes the flexible surgical tool [25, 26] despite its different postures (bending patterns) captured by real-time 3D US images based on the 3D US calibration system explained in Section 4.3 . In this system, a 3D US probe, to which a 6-DOF (degree of freedom) EM sensor is attached, is exploited to acquire the images of the surgical tool, and an EM transmitter is fixed at a certain point in the environment as the origin of the world coordinate system. From the 3D US images, the skeletons of the surgical tool with different postures are extracted, and then the coordinates in the world coordinate system of 15 labeled points along the surgical tool are calculated by the proposed 3D US calibration system.

6.1.1 Methods

(1) Overview

For current FETO surgeries, the localization of the surgical tool remains a challenging problem. Here, a block diagram of the proposed method for the accurate localization of the surgical tool is shown in Fig. 6.1.

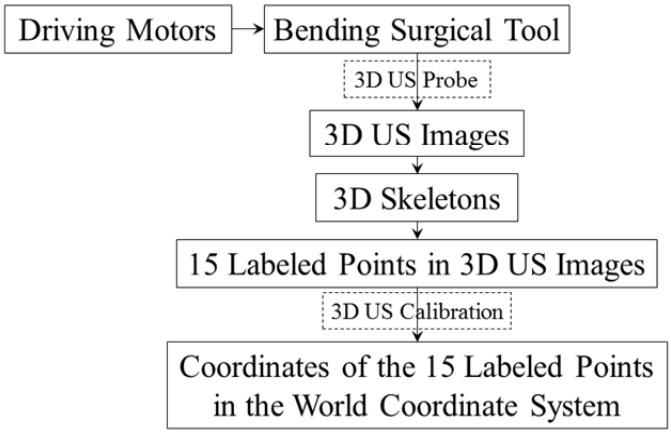
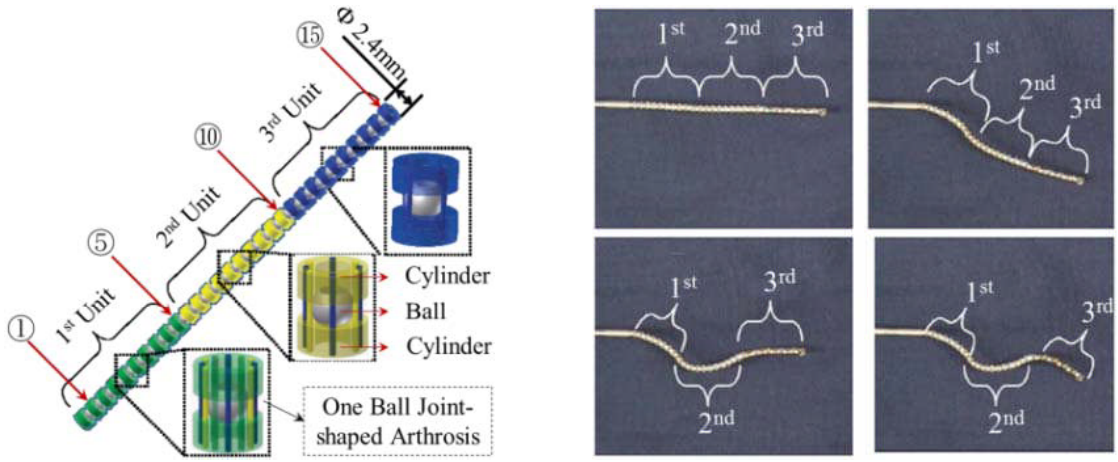


Figure 6.1 Diagram of the proposed approach.

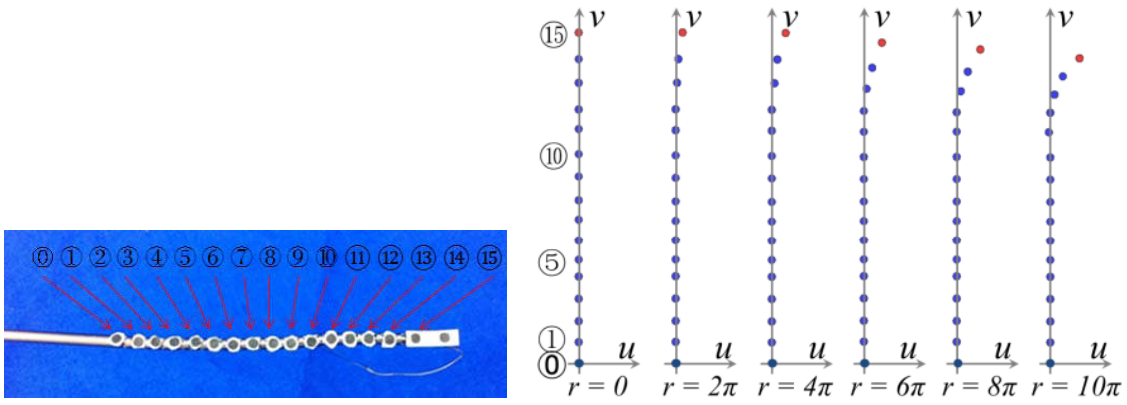
(2) Flexible Surgical Tool

Zhang et al. [25, 26] developed a wire-driven surgical tool with an outer diameter of 2.4 mm and an inner diameter of 1.0 mm so that the tool can be inserted into the trocar with a diameter of 3 mm. As shown in Fig. 6.2 (a), the surgical tool consists of three units, each of which contains 10 ball-joint-shaped arthroses. Each unit also has two-DOFs, which enable the surgical tool to bend in different bending motions (patterns), as shown in Fig. 6.2 (b).



(a) (b)

Figure 6.2 The structure and bending patterns (motion) of the surgical tool.



(a) 15 labeled points (b) The deformations of the 15 labeled points

Figure 6.3 The deformations of the 15 labeled points in the surgical tool.

In order to localize the surgical tool, 15 points are labeled as No.1 to No.15 by selecting one out of every two arthroses (points) along the surgical tool, as shown in Fig.

6.3 (a). The deformations (bending patterns) of the surgical tool (Fig. 6.3 (b)), are determined by the driving motor's rotation r (radian), and the motion data with the different bending patterns of the 15 labeled points are obtained by an optical camera (Here, a simple bending of the surgical tool in one direction is firstly considered).

(3) Proposed Approach

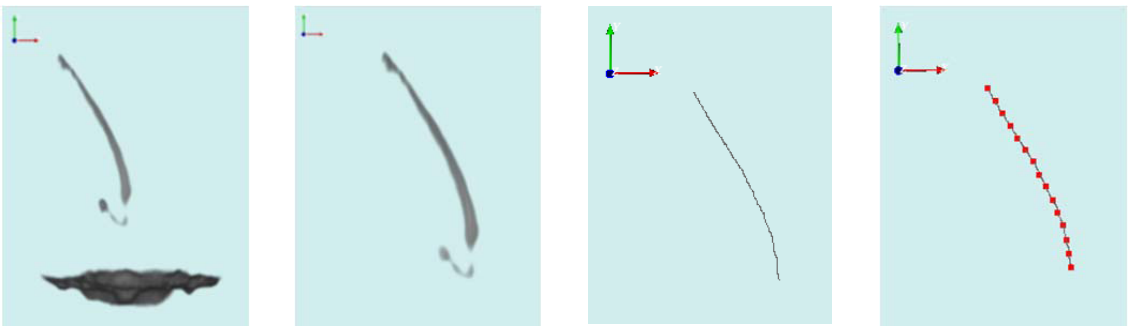
To accurately localize the surgical tool, the coordinates of the 15 labeled points along the surgical tool in the world coordinate system need to be estimated. The procedure is as follows (Fig. 6.4):

Step-1 After segmentation of the 3D US volume by global threshold, the surgical tool is detected by selecting a region of interest (ROI) based on the segmentation result, and the 3D distance map in the ROI is calculated;

Step-2 The skeleton is extracted by a fast marching minimal path extraction in ITK [153], where the start, the end, and several way-points in the skeleton are required before extraction (although these points are designated manually at present, they can be provided automatically in the future by attaching several 3D EM tracking sensors along the surgical tool);

Step-3 The coordinates of the 15 labeled points in 3D US space are measured in the extracted skeleton based on the known coordinates of the start and end points, because the distance between two adjacent labeled points is the same.

Step-4 The coordinates of the 15 labeled points in the world coordinate system are calculated by the proposed 3D US calibration system.



(a) 3D US data (b) 3D ROI (c) skeleton (d) 15 labeled points

Figure 6.4 The procedure of the surgical tool localization.

6.1.2 Results

Figure 6.5 shows the mean and standard deviation of average and maximum TREs for the 15 labeled points corresponding to each bending pattern in Fig. 6.3 (b). In Fig. 6.5, TREs denote the RMS distance errors between the coordinates of the 15 labeled points transformed from 3D US image space to 3D tracking space and the corresponding coordinates localized in 3D tracking space. For each bending pattern from $r = 0$ to $r = 10\pi$, the TREs of individual labeled points show slight differences because the standard deviations of average and maximum TREs (blue and red dashed lines in Fig. 6.5) are both smaller than 0.5 mm. The fluctuations on the means of average and maximum TREs (blue and red solid lines in Fig. 6.5) are also small. Besides, considering TREs of all points for all bending patterns, an average TRE of 1.99 ± 0.49 mm and a maximum TRE of 2.61 ± 0.49 mm are obtained, which satisfy the requirement of the FETO surgery: less than 3.85 mm.

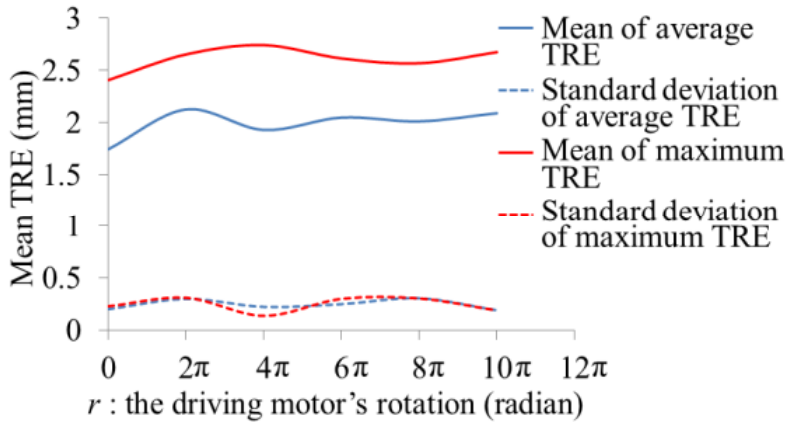


Figure 6.5 Validation of the 15 labeled points for the surgical tool localization.

In FETO surgeries, the surgical tool is inserted into fetal airways until the entrance of fetal trachea, and the narrowest part of the passage is the fetal pharynx where the surgical tool is supposed to go through. In addition, FETO surgery is operated at about 26-29 weeks gestational age (GA) [23], and the average diameter of fetal pharynx is 7.7 mm at about 26-30 weeks [154]. Thus, 3.85 mm - half of the average diameter of fetal pharynx is the maximum distance error tolerable for this surgery. Therefore, the localization of the surgical tool in the experiments achieves sufficiently high accuracy for FETO surgeries.

6.1.3 Conclusions

This section has proposed an approach that localizes the surgical tool for FETO surgeries from 3D US images based on the 3D US calibration system. The localization of the surgical tool achieves an average TRE of $1.99 \pm 0.49\ mm$ and a maximum TRE of $2.61 \pm 0.49\ mm$, which satisfy the requirement of the FETO surgery: less than 3.85 mm. The experimental results demonstrate that the proposed approach can achieve sufficiently high accuracy for localizing the surgical tool. To further improve the calibration consistency and accuracy, one potential of the future work is to design new algorithms to automatically identify the vertices in the cones' phantom and extract the skeletons of the surgical tool. Furthermore, the effect of speed of sound would be taken into consideration in the future.

6.2 Fetal Mouth Tracking

6.2.1 Introduction

To perform the FETO surgery, a fiber fetoscope with a diameter of 1.3 mm within a cannula (Karl Storz) featuring a diameter of 3.3 mm is inserted into the amniotic cavity through the abdominal and uterine walls, toward the fetal mouth and trachea, and is navigated by the ultrasound (US) images and fetoscopic images [22, 23], but this operation is so dangerous that the use of a supporting system for navigating surgeries is imperative. On the other hand, although fetoscopic images have been utilized as a tool for guidance of FETO surgeries, they just provide a direct observation for surgeons, and there is still no further processing in the fetoscopic images to guide a surgical tool to be inserted into the fetal trachea. In order to address this problem, an automatic approach is proposed to detect and track the fetal face and mouth accurately in real-time in a fetoscopic video sequence, as well as provide an effective and timely feedback to the robot control system of the surgical tool.

6.2.2 Methods

(1) Mains Steps of the Proposed Approach

To detect and track the fetal face and mouth in a fetoscopic video sequence, there are four main steps as follows. Figure 6.6 shows the block diagram of the method.

Step-1 The region of interest (ROI) for each frame in the fetoscopic video sequence is selected based on the mask image, and smoothed by Gaussian filter;

Step-2 The fetal face in ROI is detected by the AdaBoost algorithm based on Haar-like features [155];

Step-3 The fetal face is tracked by the CamShift algorithm based on an iterative search in a color histogram [156];

Step-4 In the search window from the CamShift algorithm, the area of the fetal mouth is segmented by a global threshold, and the edges of the fetal mouth are extracted by canny edge detection. After this, an ellipse is fitted to the edges of the fetal mouth by the improved iterative randomized Hough transform method proposed in [118, 119] and Section 4.2 ;

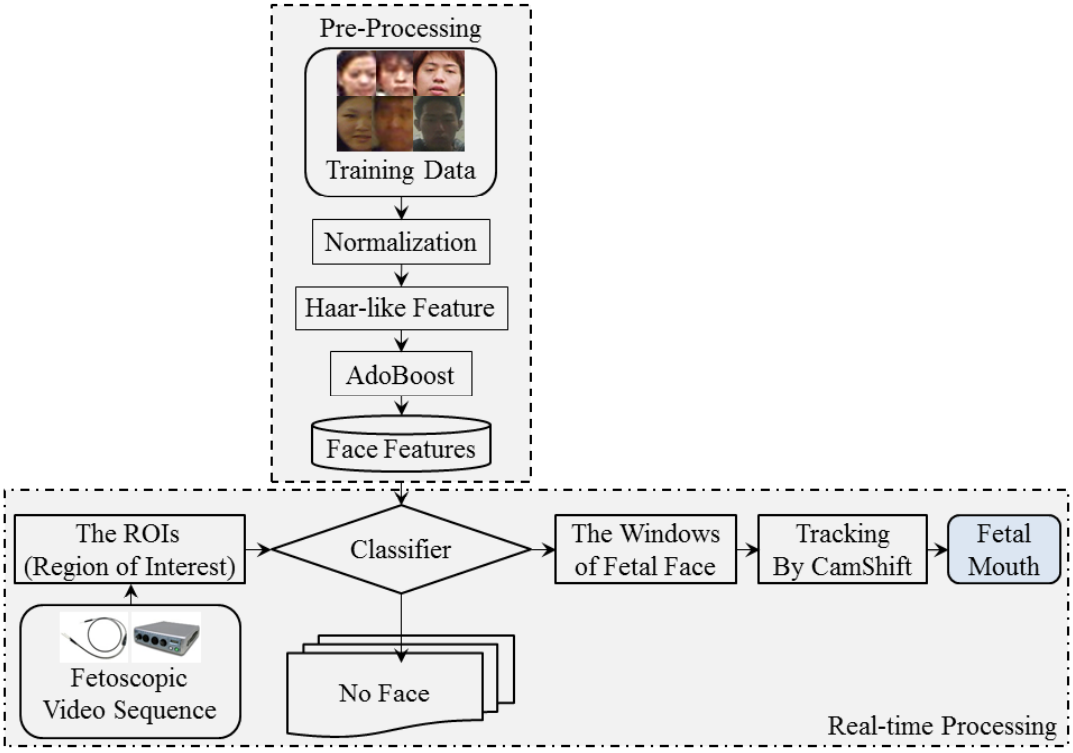
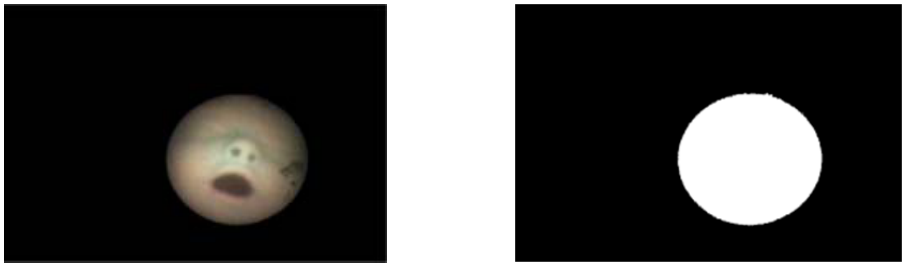


Figure 6.6 The block diagram of the proposed approach.

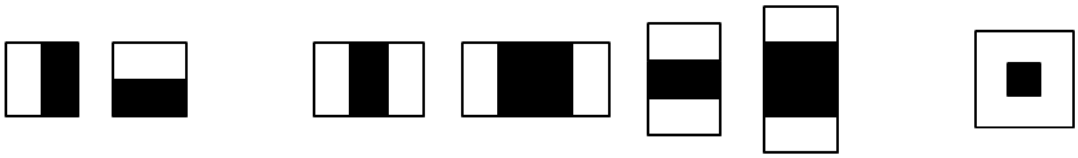
(2) Region of Interest

The frames in the fetoscopic video sequence have the same ROI. Thus, a mask image is used to select the ROI for each frame, i.e., the white area in Fig. 6.7 (b).



(a) one frame (b) the mask image

Figure 6.7 One frame and the mask image.



(a) (b) (c)

Figure 6.8 Haar-like features.

(3) Classifier

(3-1) Haar-like Features

As shown in Fig. 6.8, Haar-like features, such as (a) edge features; (b) line features; and, (c) center-surround features, are often used for human face detection. The Haar-like features are calculated by scanning all of the pixels from left to right and top to bottom in each frame. However, the direct calculation of the pixels' sums over multiple rectangles would make the detection task very slow and unable to be realized in real-time. An efficient method was proposed in [155] for computing the sums quickly by using an intermediate representation for the image, which is called as the integral image. Then, the integral image, Summed Area Table (*SAT*), is computed over the whole image *I*, and the *SAT* is defined as

$$SAT(x,y) = \sum_{i<x, j<y} I(i,j) \tag{6.1}$$

The pixel sum over a generic rectangle $r = \{(x,y), x_0 \le x < x_0 + w, y_0 \le y < y_0 + h\}$ can then be computed using the *SAT*. The sum is computed by using the rectangle corners coordinates, as indicated in Eq. (6.2).

$$S(r) = SAT(x_0 + w, y_0 + h) - SAT(x_0 + w, y_0) - SAT(x_0, y_0 + h) + SAT(x_0, y_0) \tag{6.2}$$

(3-2) AdaBoost Algorithm

AdaBoost is a machine learning algorithm that selects a small number of critical visual features from a larger set and yields an extremely efficient classifier [155, 157]. Originally, the AdaBoost algorithm was used to boost the classification performance of a simple (or weak) learning algorithm. The weak learning algorithm is designed to select the single rectangle feature that best separates the positive and negative examples. For each feature, the weak learner determines the optimal threshold classification function, such that the minimum number of examples is misclassified. A weak classifier $h_j(x)$ thus consists of a feature f_j, a threshold θ_j, and a parity p_j indicating the direction of the inequality sign:

$$h_j(x) = \begin{cases} 1 & if \quad p_j f_j(x) < p_j \theta_j \\ 0 & otherwise \end{cases} \tag{6.3}$$

where x is a 24×24 pixel sub-window of an image, and the summary of the boosting process is shown in Fig. 6.9. Then, the final strong classifier $H(x)$ is a linear combination of T weak classifiers $h_j(x)$:

$$H(x) = sign(\sum_{j=1}^{T} \alpha_j h_j(x)) \tag{6.4}$$

where α_j is the weight assigned for each weak classifier $h_j(x)$.

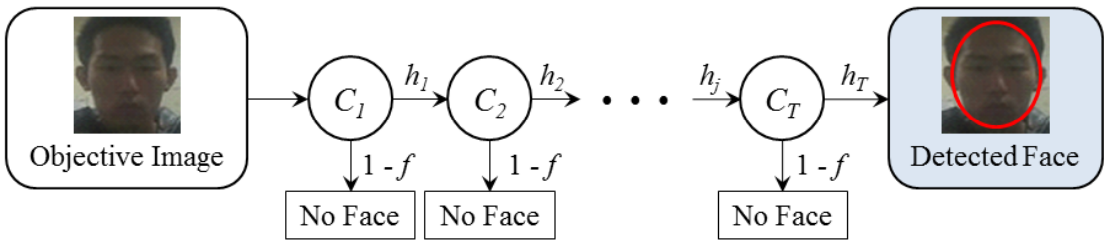


Figure 6.9 The boosting process of the AdaBoost algorithm.

(3-3) CamShift Algorithm

The continuously adaptive mean shift algorithm (CamShift) [5] is an adaptation of the mean shift algorithm for object tracking that is intended as a step toward head and face tracking for a perceptual user interface. Figure 6.10 summarizes the algorithm described below. For each video frame, the raw image is converted to a color probability

distribution image by a color histogram model of the color being tracked. The center and size of the color object are found via the CamShift algorithm operating on the color probability image. The current size and location of the tracked object are reported and used to set the size and location of the search window in the next video image. The process is then repeated for continuous tracking. Figure 6.11 gives an example of the tracking of fetal facial color by color histogram and color distribution.

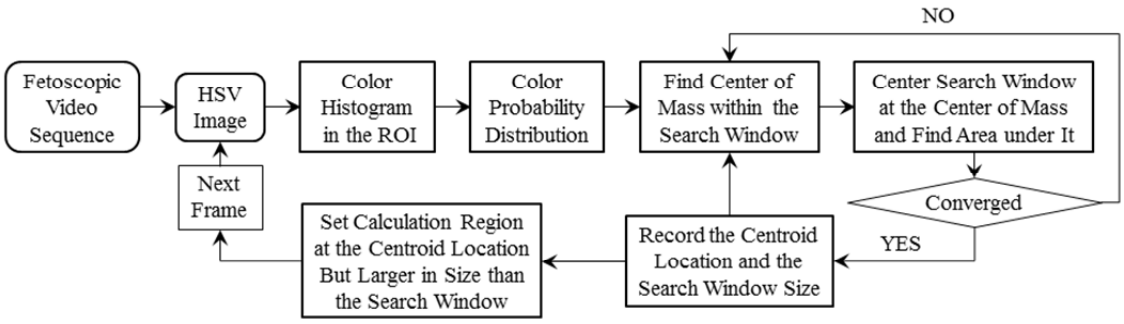


Figure 6.10 The flow diagram of color object tracking.

(a) one frame (b) color histogram (c) color distribution

Figure 6.11 Fetal facial color tracking.

(3-4) Fetal Mouth Detection

In order to locate the fetal mouth in the search window from the CamShift algorithm after tracking the fetal face successfully, the improved iterative randomized Hough transform (IRHT) method proposed in [118, 119] and Section 4.2 is exploited to improve the efficiency and accuracy of detecting the ellipse of the fetal mouth for real-time tracking. Figure 6.12 illustrates the process of the fetal mouth (as an ellipse) detection: (a) one frame with the fetal mouth; (b) the edges detected from the ROI of the frame in (a); and, (c) the ellipse detected from the edges in (b).

On the other hand, the fetoscope should be relatively close to the target, and then, the target will be easily distinguished. Accordingly, the change of the target size would

not be great in a fetoscopic video sequence when the fetoscope is close to the target. As a result, the size of the ellipse is restricted to be fitted to the fetal mouth in the range of $20 \leq a \leq 40, 10 \leq b \leq 25$, and $0.5 \leq e \leq 0.9$ based on the used fetal phantom so that the efficiency and accuracy of the improved IRHT algorithm are improved, where a and b denote the major and minor semi-axes of the ellipse, respectively, and e denotes the eccentricity of the ellipse. Certainly, for other fetal phantoms or real fetuses of different sizes, specific parameter values of the restriction should be trained and predefined. At the same time, once the fetal mouth is detected from one frame, then the ROI estimated from the detected ellipse (the gray rectangle in Fig. 6.12 (c)) will be utilized in the next frame to reject the edges that do not belong to the fetal mouth for a better tracking efficiency, where the center of the rectangle ROI is the same as the detected ellipse, and the width and height are defined as $a \times 2.6$ and $b \times 2.6$, respectively.

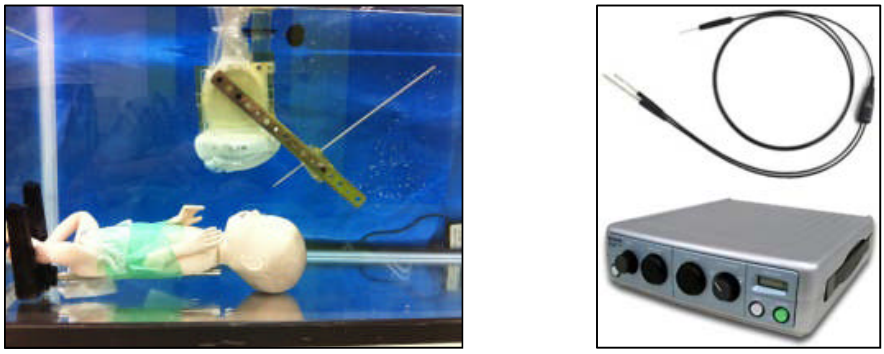
(a) (b) (c)

Figure 6.12 Fetal mouth detection.

6.2.3 Results

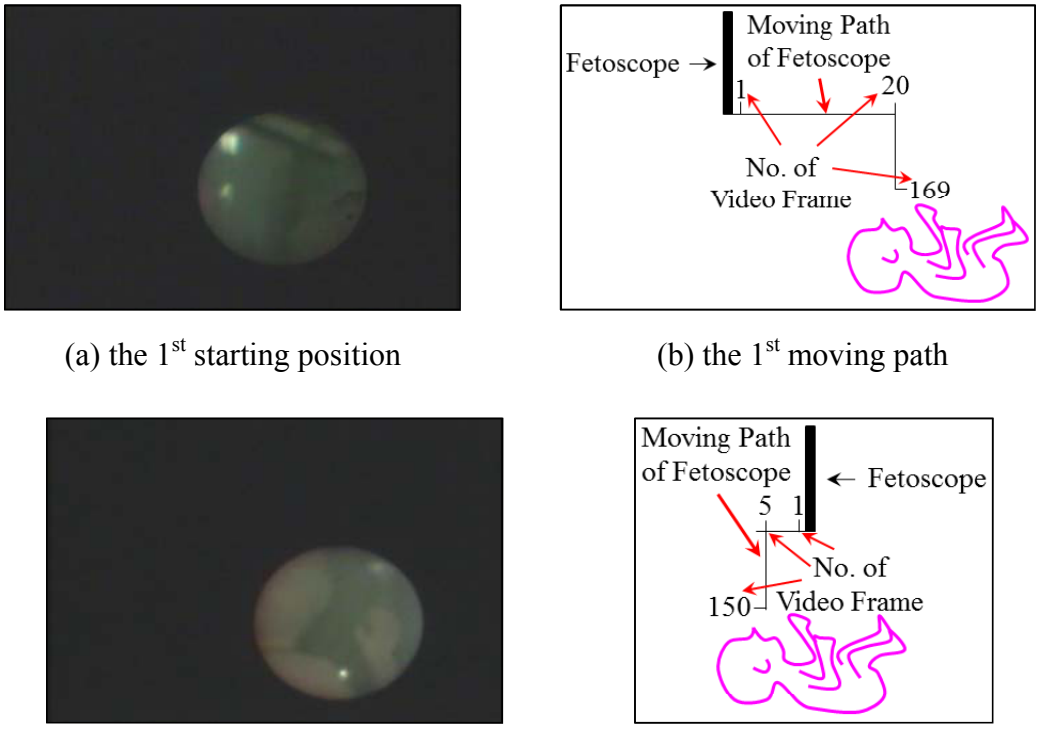
In the experiments, as shown in Fig. 6.13 (a), a fetal phantom was placed on the bottom of a water tank, and a 3D ultrasound probe was fixed above the phantom by a plastic holder, whose beam surface contacted the water. A straight cannula was exploited to simulate the movement of the surgical tool as a real operation, and a fetoscope, which is a fiber soft endoscope (FiberTech) with a high resolution CCD camera based on a FT-201F imaging system (in Fig. 6.13 (b)), was inserted into the cannula to obtain the video sequence. The fetoscope was moved from two different starting positions toward the position of the fetal mouth. As shown in Fig. 6.14, (a) is the 1st starting position from which the fetus cannot be viewed by fetoscope; (b) is the corresponding moving path of the fetoscope from the 1st starting position; (c) is the 2nd

starting point from which the fetal left hand can be viewed by the fetoscope; and, (d) is the corresponding moving path of the fetoscope from the 2nd starting position.



(a) (b)

Figure 6.13 Experimental equipment.



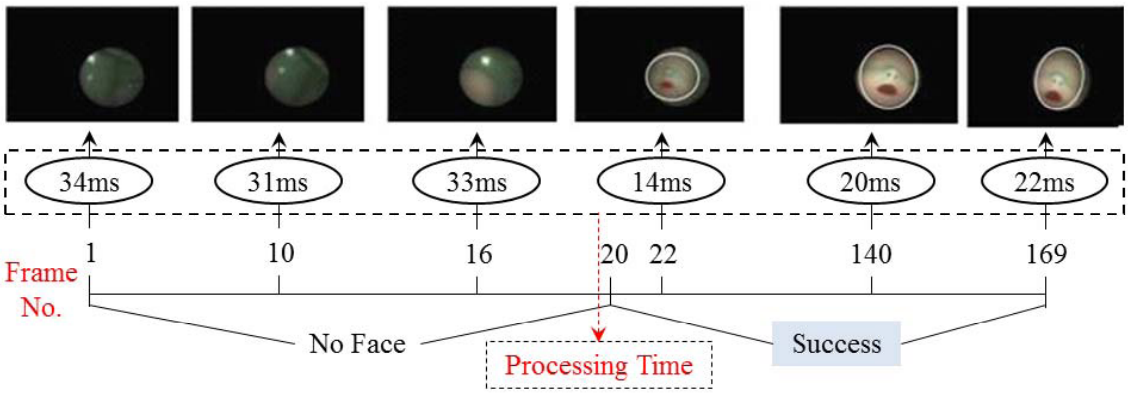
(a) the 1st starting position (b) the 1st moving path

(c) the 2nd starting position (d) the 2nd moving path

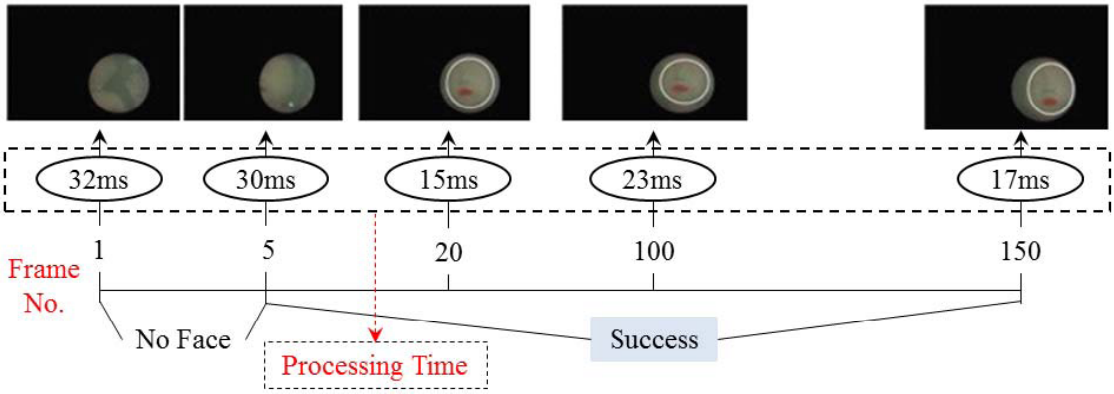
Figure 6.14 Two different starting positions and the corresponding moving paths.

In Fig. 6.15, (a) and (b) show the first and second process of tracking of the fetal face and mouth from the 1st and 2nd starting positions, respectively. The fetal face was tracked successfully and the mouth was detected accurately from frame No. 20 to No. 169 for the 1st process and from frame No. 5 to No. 150 for the 2nd process, respectively.

In the 1st process, the average processing times before and after the successful face tracking are 32.5 ± 4.88 ms and 20.9 ± 6.48 ms, respectively. In the 2nd process, the average processing times before and after the successful face tracking are 30.7 ± 3.57 ms and 23.8 ± 5.59 ms, respectively, as shown in the columns of "No Face" and "Face Tracking 1" in Table 6.1. Considering the frame rate of 29 FPS of the FT-201F imaging system, if the processing time for each frame is less than 34.5 ms, then the real-time tracking can be realized. Thus, from the average processing times of the 1st process and 2nd process, the proposed approach basically achieves the real-time tracking of the fetal face and mouth in the fetoscopic video sequence.



(a) the 1st process about the tracking of fetal face and mouth from the 1st starting position



(b) the 2nd process about the tracking of fetal face and mouth from the 2nd starting position

Figure 6.15 The processes of the tracking of the fetal face and mouth.

Additionally, a comparison between the improved IRHT method and the original IRHT method is made for fetal mouth detection. In Table 6.1, the average processing times for the tracking of the fetal face and mouth by the improved IRHT method, which

is shown in the column of "Face Tracking 1," are significantly smaller than those values by the original IRHT method shown in the column of "Face Tracking 2". In Table 6.2, the number of pixels on the detected ellipse by the improved IRHT method is quite larger than those values by the original IRHT method. Furthermore, the average processing time by the original IRHT method after the successful face tracking for the 1st process and 2nd process are 53.6 ± 11.38 ms and 58.3 ± 9.75 ms, respectively, as shown in the column of "Face Tracking 2" in Table 6.1, which cannot satisfy the requirement of real-time tracking. In conclusion, the improved IRHT method performs a better efficiency and accuracy of the fetal mouth detection, which contributes to the real-time tracking of the whole process.

Table 6.1 The average processing times for different stages and different methods (ms).

	No Face	**Face Tracking 1**	**Face Tracking 2**
The 1st Process	$32.5 \pm 4.88\ ms$	$20.9 \pm 6.48\ mm$	$53.6 \pm 11.38\ mm$
The 2nd Process	$30.7 \pm 3.57\ ms$	$23.8 \pm 5.59\ ms$	$58.3 \pm 9.75 ms$

Table 6.2 The number of pixels on the detected ellipse for two methods.

	Face Tracking 1	**Face Tracking 2**
The 1st Process	$76 \pm 8\ pixels$	$46 \pm 15\ pixels$
The 2nd Process	$81 \pm 7\ pixels$	$39 \pm 12\ pixels$

6.2.4 Conclusions

This section has proposed an approach that automatically detects and tracks the fetal face and mouth during the fetoscopic video sequence for navigating a surgical tool to be inserted into the fetal trachea. Firstly, the AdaBoost algorithm is utilized as a classifier to detect the fetal face, based on Haar-like features. After that, the CamShift algorithm is applied to track the fetal face based on the image color histogram, and the fetal mouth is detected as an ellipse by the improved IRHT method. The experimental results demonstrate that the proposed approach can automatically and accurately detect and track the fetal face and mouth in real-time in the fetoscopic video sequence, as well as can provide an effective and timely feedback to the robot control system of the

surgical tool. In the future, the real-time tracking of the fetal face and mouth in the fetoscopic video sequence will be combined with the distance estimation between the tip of the surgical tool and the fetal mouth from 3D ultrasound data, as this can provide more accurate and reliable feedback.

Chapter 7 Estimating the Position of the Fetal Trachea

7.1 Introduction

Real-time 3D US images as an intraoperative guide have been introduced to fetal MIS surgeries [32, 33], due to their cost efficiency, real-time imaging, lack of radiation exposure, and impact-free characteristic compared to preoperative MR or CT modalities. Unfortunately, 3D US images have many disadvantages. The imaging quality is very poor because of the speckle noise. The imaging field of view (FOV) is also quite limited. Often, some critical internal organs are difficult to visualize and identify from 3D US volumes due to their low resolution and the echogenicity relative to the target organs.

For the FETO surgery, real-time 3D US images provide a helpful guidance during the operation. However, the critical internal organs, such as fetal trachea, that is easily to be hurt by the surgical tool, cannot be identified from 3D US images. Thus, in order to estimate the position of the fetal trachea, an approach is proposed in this chapter.

7.1.1 Proposed Approach

For the FETO surgery, the location of the surgical tool, the fetus and the critical internal organs plays an important role on reducing the surgical risk. On one hand, 3D US images can allow only 3D modeling of the surface of a fetus, but cannot 3D-model internal organs of a fetus, because the internal organs cannot be distinguished in 3D US images. Although MR images are available to 3D-model internal organs of a fetus in actual clinical applications, in case of the research in this thesis, it is very difficult to get a permission to use real fetal MR images due to privacy problems in Japan. Therefore, as an alternative for MR images, models of fetal phantoms with internal structure are used. The model is registered to the 3D US images so that the internal organs are located in the 3D US images.

To do this, resin fetal phantoms without internal structure are used instead of real fetus in the mother's body, where normally resin phantoms with internal structure are

also not available. Since the resin phantom does not have the internal structure, after the phantom is laser-scanned, the phantom's 3D model is reconstructed, and the internal structure in the phantom's model is designed. Accordingly, a surgical navigation scheme is proposed based on fetal phantoms, the designed internal structure, real-time 3D US images, and the novel 3D US calibration system proposed in Section 4.3 .

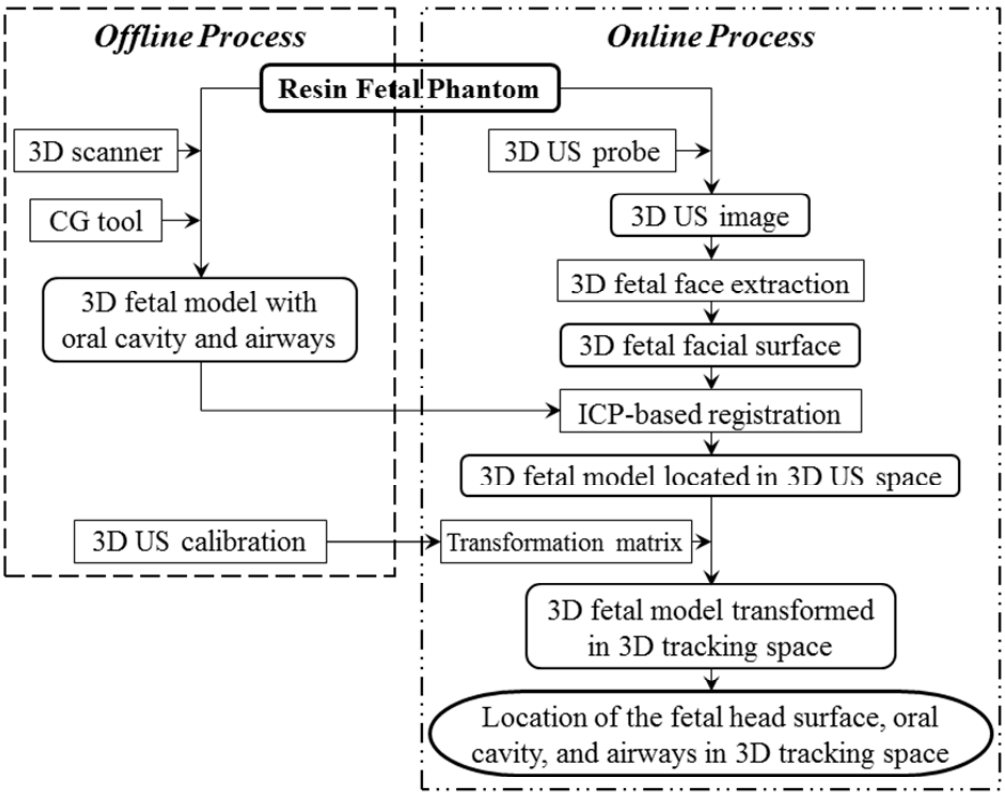


Figure 7.1 The block diagram of the proposed approach.

Figure 7.1 shows a block diagram of the proposed approach, which consists of an offline and online process. The offline process prior to surgery includes reconstruction of 3D fetal modal with oral cavity and airways, and estimation of the transformation matrix of the 3D US calibration system by the novel cones' phantom and EM tracking device. The online process starts with obtaining 3D fetal US images by scanning a fetal phantom in a water tank with a 3D US probe. Then, a 3D fetal facial surface is extracted from the 3D US images and registered with the 3D fetal model. After the registration, the 3D fetal model located in the 3D US space is transformed into the 3D tracking space by the estimated transformation matrix, where, as explained in previous sections, the 3D US space denotes the coordinate system of 3D US images, and the 3D tracking space denotes the world coordinate system of the 3D EM transmitter. Finally, the 3D fetal facial surface, oral cavity, and airways are accurately located in the 3D tracking space.

7.2 3D Fetal Model

7.2.1 3D Fetal Facial Surface

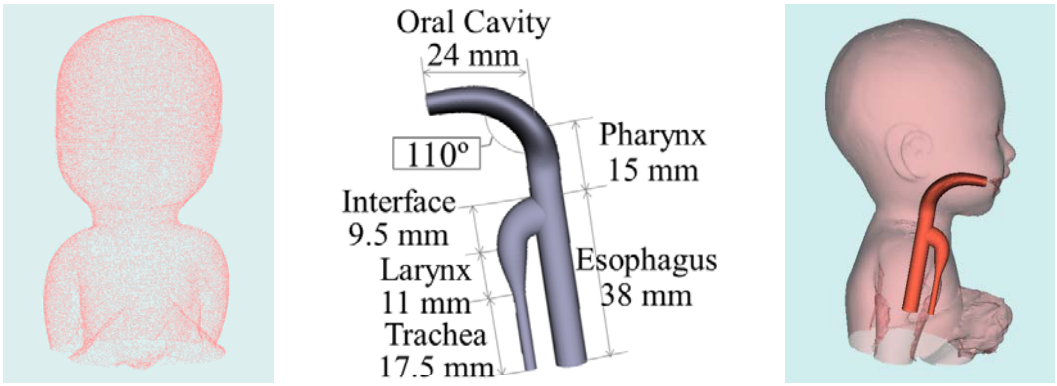
Since the internal anatomies of the oral cavity and airways are difficult to visualize and distinguish from 3D US volumes due to their low resolution, a 3D fetal model with such anatomies is designed. At first, the surface of the 3D fetal model is created by scanning a physical fetal phantom with an untouched high-speed 3D scanner - 3030RGB/MS, as shown in Fig. 7.2. First, the phantom is made to stand in the front of the scanner. Next, the phantom on the rotating platform rotates along a complete circle, and each turn is set to 10°. Finally, the 3D fetal facial surface can be reconstructed as 3D point cloud data obtained by the untouched 3D scanner, as shown in Fig. 7.3 (a).

Figure 7.2 Untouched 3D scanner and the standing fetal phantom.

7.2.2 Design of the Fetal Oral Cavity and Airways

The designed internal anatomies include the structures of the oral cavity, pharynx, larynx, trachea, and esophagus. As the FETO surgery is often performed at about 26-29 weeks gestational age (GA) [23], the average size for each part is selected at about 24-32 weeks GA. These structures [158] are drawn as shown in Fig. 7.3 (b) by a 3D CAD design software (Solidworks 2008). To simplify the design, the cross-sectional shapes of all tracts are considered as ellipses or circles. The size of each part is

determined by referring to [154, 159, 160]. Through the manual rotation and translation of the designed structure in 3D space, a 3D fetal model with an oral cavity and airways can be achieved as shown in Fig. 7.3 (c).



(a) (b) (c)

Figure 7.3 3D fetal model. (a) 3D point cloud, (b) the design of the fetal oral cavity and airways, (c) 3D fetal model.

7.3 3D Rigid Registration

7.3.1 3D Fetal Ultrasound Images

In the online process, a physical fetal phantom is scanned in a water tank with the Philips iU22 US system and a V6-2 3D US probe to obtain 3D US images (Fig. 7.5 (a)). The V6-2 3D US probe has a 2 to 6 MHz extended operating frequency range and a field of view of 66°. All acquired volumes have a size of $512(x) \times 510(y) \times 256(z)$ voxels with the scaling factors $s_x = 0.35\, mm/voxel$, $s_y = 0.23\, mm/voxel$, $s_z = 0.43\, mm/voxel$, where the $x - y - z$ coordinate system of the 3D US image, which is decided by the position and orientation of the 3D US probe, is defined in Fig. 7.5 (a).

As shown in Fig. 7.4, in order to simulate real fetus in human amniotic fluid, a fetal phantom is placed in a water tank, and then a 3D US probe is placed above the fetal head by a customized holder, to make the fetal head imaging in the center of 3D US space, since the fetus is assumed to be immobilized in uterus during the operation. Simultaneously, an EM tracking sensor is fixed on the customized holder, and the EM transmitter is laid near the water tank.

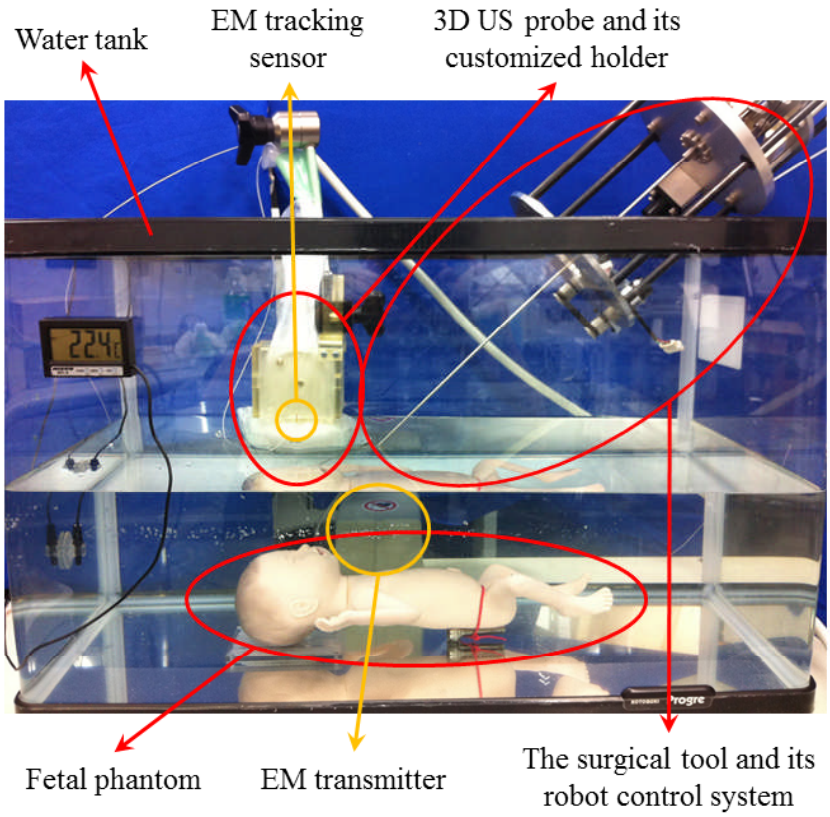


Figure 7.4 Experimental environment.

The phantom based experimental environment should be able to simulate the environment of the real fetus in the uterus (amniotic fluid) accurately. In order to compare difference of ultrasound imaging between the experimental environment and the environment of the real fetus, sound speeds in and densities of water and the used phantoms were measured at room temperature as indicated in Table 7.1, in which sound speeds in and densities of amniotic fluid and water are also listed according to [161]. Note that sound speeds in, densities of, and acoustic impedances (Z_{ai}) of water and amniotic fluid are similar to each other, where acoustic impedance Z_{ai} is calculated by

$$Z_{ai} = \rho D \tag{7.1}$$

In Eq. (7.1), ρ and D denote the density of and sound speed in water, respectively. Table 7.1 shows that the difference between the sound speeds measured in this thesis and shown in [161] in water at room temperature is approximately 0.9%, which is very small. Thus, it can be inferred that the error between the measured sound speed in the used fetal phantoms and their ground truth should be acceptably small. By Eq. (7.1), the acoustic impedance of the used phantoms is calculated as 1.16, as shown in Table 7.1.

Table 7.1 Sound speeds, densities and acoustic impedances of different materials.

Materials	**Sound speed (D)** ($m\ s^{-1}$)	**Density (ρ)** ($10^3 kg\ m^{-3}$)	**Acoustic impedance (Z_{ai})** ($10^6 kg\ m^{-2}\ sec^{-1}$)
amniotic fluid (ref. [161])	1510	1.01	1.53
water (ref. [161])	1500	1	1.50
water (measured)	1486	1	1.49
fetal phantoms (measured)	1208	0.96	1.16
human skin (ref. [162])			1.68
human muscle (ref. [163])			1.70
human skull (ref. [163])			6.10

Table 7.2 The values of R_{US}(%) at various boundaries.

Boundary	**The reflection coefficient (R_{US})**
amniotic fluid – human skin	0.2%
human skin – human muscle	0.004%
human muscle – human skull	31.8%
water (measured) – fetal phantoms	1.6%

For ultrasound imaging, ultrasonic waves are emitted from a transducer, and partly reflected back to the transducer as echoes from boundaries between two adjacent objects. The amount of the reflected echoes depends on the difference in acoustic impedances between the two objects such as tissues traversed by the beam, where this difference in Z_{ai} is commonly referred to as impedance mismatch. The larger the impedance mismatch is, the larger the amount of reflection is, and the smaller the amount of transmission is. The reflection coefficient R_{US} is defined as a ratio of the intensity of the reflected wave relative to that of the incident wave as follows [163],

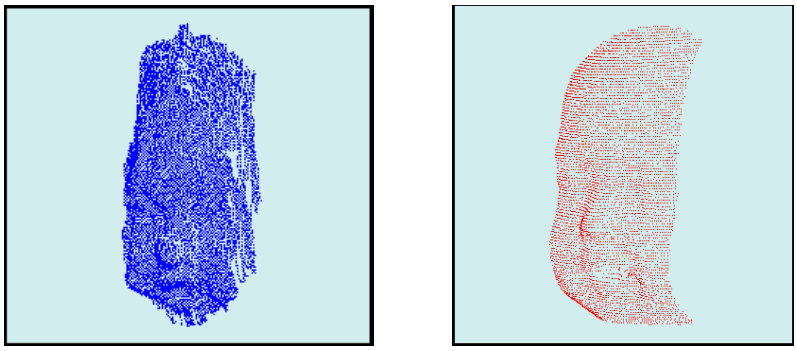
$$R_{US} = \frac{(Z_{ai2} - Z_{ai1})^2}{(Z_{ai2} + Z_{ai1})^2} \tag{7.2}$$

where Z_{ai1} and Z_{ai2} are the acoustic impedances of the two objects that form the boundary. Table 7.2 shows the values of R_{US} (%) at various boundaries. At boundary between amniotic fluid and human skin, only 0.2% of the incident energy is reflected, and 99.8% of the energy is transmitted across the boundary. Although this reflected energy is much smaller than 31.8%, the reflection coefficient at boundary between human muscle and human skull, it is sufficient to reveal the borders of fetal faces according to [90], in which clear clinical 3D US data of fetal faces are presented. Thus, the faces of the used fetal phantoms in a water tank can be imaged reasonably due to the reflection coefficient value 1.6% at boundary between water and the phantoms. Then, the gray-levels of the phantoms' faces should be between the gray-levels for real fetal faces and real fetal skulls in 3D US images, while gray-levels for the phantoms and real fetal faces could be close to each other. Actually, by visually comparing US images for the used phantoms with real fetal faces [90] and skulls [164], it turns out that gray-levels for the used phantoms and real fetal faces are almost identical, while real fetal skulls are brighter due to the larger reflection coefficient value (31.8%).

7.3.2 3D Fetal Facial Surface Extraction

Unlike MR or CT imaging technology, 3D US images do not have high resolution, and the imaging quality of US is easily affected by phantom material and air bubbles in the water. Therefore, a method that extracts the 3D fetal facial surface by scanning the slices in an $x-y$ plane is proposed. Herein, a fetal phantom at the gestational age of 30 weeks is used as an example, and its 3D US data (Fig. 7.5 (b)) that has been pre-processed by Gaussian filter and threshold segmentation are shown in Fig. 7.5 (c), where 3D US probe is placed above the fetal head, and the fetal head can be centered in the 3D US space ($x-y-z$ coordinate system), because the fetus is assumed to be fixed in uterus during the operation. The procedure for this method is described in Section 5.2.2, and the 3D point cloud of the fetal facial surface can be extracted from 3D US images (Fig. 7.5 (c)), as shown in Fig. 7.5 (d). On the other hand, it is straight forward to extract the corresponding facial surface from the 3D fetal model simply by setting the range on the *x*, *y*, and *z* axes (Figure 7.5 (e)).

(a) 3D US probe (b) original 3D US (c) extracted 3D US



(d) 3D facial surface (e) another surface

Figure 7.5 3D fetal facial surface extraction from 3D US images.

7.3.3 ICP-based Registration

The iterative closest point (ICP) method for the registration of 3D point clouds was proposed by Yang et al. [102] and Besl et al. [94]. It has been widely used in a variety of fields including medical images, because of its good accuracy and fast speed. Given two point clouds $A = \{a_i\}$ ($i = 1,2,\ldots,N_a$) and $B = \{b_j\}$ ($j = 1,2,\ldots,N_b$), the transformation between data A and data B is assumed to be linear with respect to a rotation matrix R and translation vector t. The goal of the ICP algorithm is to find the transformation parameters for minimizing the error (mostly least squares) between the transformed data B and the closest points of the data A, which is described by Eq. (7.3) [165], where $a_{c(i)}$ is the point of the data A paired with b_i.

$$(R,t) = \arg\min_{R,t}\left(\sum_{i=1}^{N_b} \| a_{c(i)} - (Rb_i + t) \|^2\right) \tag{7.3}$$

(1) Coarse Registration

For the purpose of improving the efficiency, accuracy, and robustness of the conventional ICP algorithm, an ICP-based registration algorithm is proposed, which consists of coarse registration and fine registration. In the coarse registration facial feature points, i.e., nose tip, two eyes' inner corners, and nose's upper bridge, are detected in the two extracted 3D fetal facial surfaces from 3D US images and the 3D fetal model. As introduced in [80], the shapes of the nose tip, two eyes' inner corners, and nose's upper bridge are respectively elliptical convex, elliptical concave, and hyperbolic convex, which can be described by the mean (H) and Gaussian (K) curvatures. From [166], a surface that has either a peak or a pit shape has a positive Gaussian curvature value ($K > 0$). The nose tip has a peak surface type, and each of the two eyes' inner corners has a pit surface type that is detectable based on the Gaussian curvature. These points should have the highest positive Gaussian curvature values among the points on the facial surface. However, since the unpredictable sharp spike noise on the 3D fetal facial surfaces may result in a false feature detection, a boosting traversal scheme based on the spatial relations among key features proposed in [146] is applied to guarantee a successful detection of the key facial features.

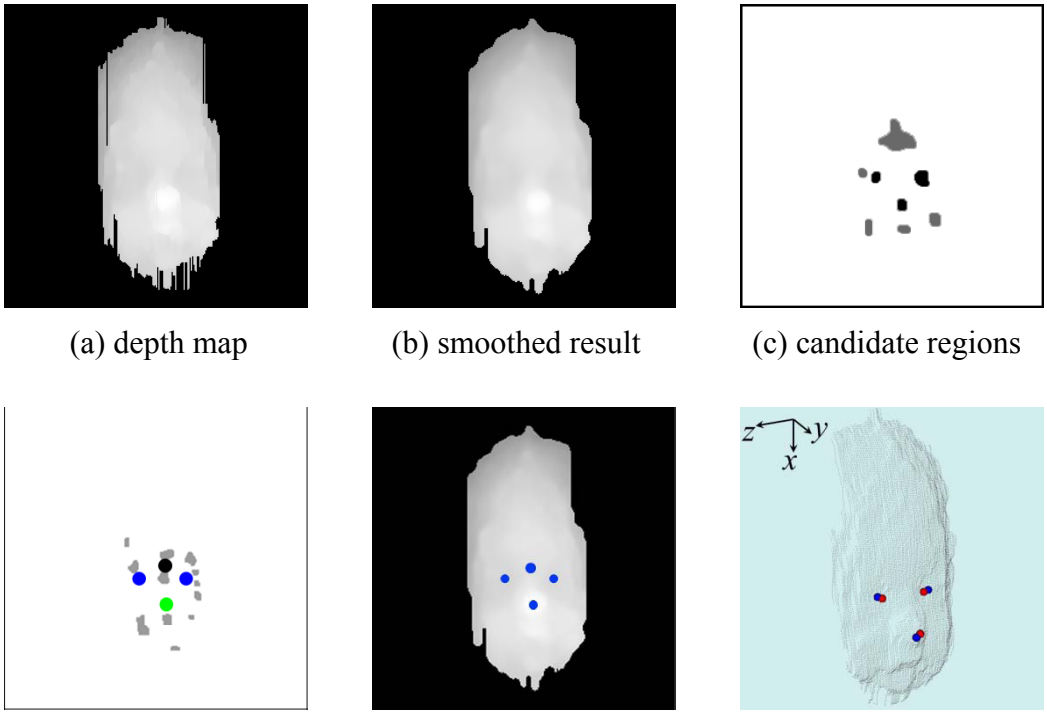
About the feature detection, the extracted 3D point cloud (Fig. 7.5 (d)) from 3D US images is used as an example for explaining the procedure. The steps are as follows:

Step-1 The depth image of the fetal facial surface is mapped to 2D space, as shown in Fig. 7.6 (a).

Step-2 The depth map is smoothed (Fig. 7.6 (b)) by a median filter (by a 5×5 window) and a Gaussian filter (by a 5×5 window, and the Gaussian sigma, which is the standard deviation of the Gaussian distribution, is set as 1.25)} to remove sharp spikes and noise. Then the depth map is thresholded to detect regions of high curvatures, and to reject points with low curvatures by $|H| \geq T_h$ and $|K| \geq T_k$, where $T_h = 0.02$ and $T_k = 0.00025$ are predefined by experiments.

Step-3 The candidate regions of the nose ($H > 0 \,\&\, K > 0$) and two eyes' inner corners ($H < 0 \,\&\, K > 0$) are selected from the thresholded depth map (Fig. 7.6 (c)), and are sorted in descending order by the standard deviation of Gaussian curvatures of each region, respectively.

Step-4 The three regions are selected in sequence from the sorted candidate regions of the nose and eyes' inner corner (e.g., three black areas in Fig. 7.6 (c)).



(a) depth map (b) smoothed result (c) candidate regions

(d) nose upper bridge (e) detected features (f) coarse registration

Figure 7.6 Fetal facial feature detection and coarse registration.

Step-5 The positions of the points with maximum Gaussian curvature in the selected three regions are labeled as feature points (Fig. 7.6 (d)) [167], where the two blue points indicate the positions of the two eyes' inner corners, and the green point indicates the position of the nose tip.

Step-6 The candidate regions for the nose's upper bridge are classified by HK classification ($H > 0, K < 0$) and the threshold of the depth map ($|H| \geq 0.005, |K| \geq 0.00005$), e.g., the gray regions in Fig. 7.6 (d).

Step-7 The candidate region closest to the midpoint of the two eyes' inner corners is selected as the region of the nose's upper bridge, whose centroid is detected as the position of the nose's upper bridge, e.g., the black point in Fig. 7.6 (d).

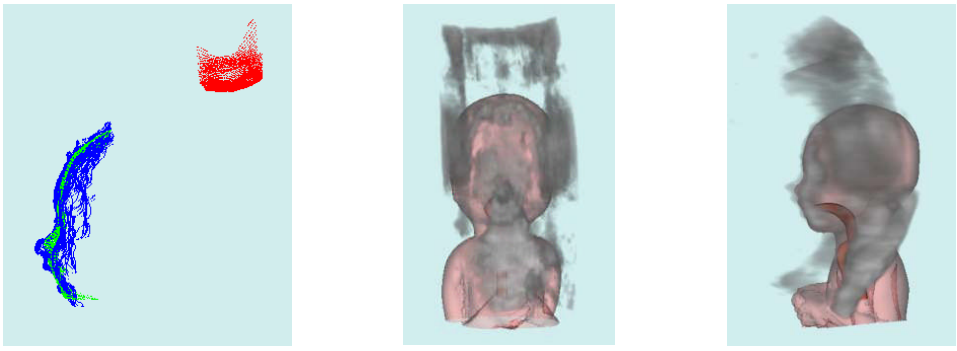
Step-8 The spatial relationship among these four features is examined by the filter proposed in [146]. If the filter is satisfied, these feature points are finally determined; otherwise, return (4) to traverse all candidate regions until the filter is satisfied.

The boosting traversal scheme based on the spatial relations among key features [146] ensures that the four feature points can be detected successfully, as shown in Fig. 7.6 (e). Similarly, the corresponding feature points of the 3D fetal model (i.e., the red

points in Fig. 7.6 (f)) are detected and registered with the four feature points detected from 3D US images (i.e., the blue points in Fig. 7.6 (f)) by the ICP algorithm, where the 3D gray point cloud in Fig. 7.6 (f) shows the 3D facial surface extracted from the 3D US images.

(2) Fine Registration

After the feature point-based coarse registration described in Sub-section (1), fine registration for the two point clouds is performed for better efficiency, accuracy, and robustness. Then the ICP-transformation matrix in the left side of Eq. (7.3)7.3 is calculated by the ICP algorithm. The result is shown in Fig. 7.7 (a), where the right red source point cloud (containing 5049 points) is extracted from the 3D fetal model, the left blue target point cloud (containing 17863 points) is extracted from the 3D US images, and the left green resulting point cloud is transformed from the right red source point cloud by the ICP-transformation matrix.



(a) (b) (c)

Figure 7.7 The rigid registration of two point clouds.

Based on the ICP-transformation matrix, all the points of the 3D fetal model with the oral cavity and airways can be transformed into 3D US space. The results in Fig. 7.7 (b) and (c) illustrate that the transformed 3D fetal model coincides very well with the fetal facial surface in the 3D US images.

7.4 Results

7.4.1 Validation for Locating the 3D Fetal Model

In experiments, five fetal phantoms with different sizes made by Kyoto Kagaku

Co., Ltd, Japan are used, and they are respectively at the gestational age of 24-weeks (Fetus1), 26-weeks (Fetus2), 28-weeks (Fetus3), 30-weeks (Fetus4) and 32-weeks (Fetus5). A testing database of 60 3D US images are collected by utilizing such five fetal phantoms. For the ICP-registration between the fetal facial surface extracted from the 3D fetal model and 3D US data, a successful detection rate of 100% is achieved for detecting four key facial features in the coarse registration based on such a testing database, and the achieved mean RMS error of the registration results is $1.04 \pm 0.17\ mm$ based on the estimated transformation matrix in Section 4.3.1.

(1) Validation for Locating 3D Fetal Facial Surface

In order to validate location accuracy of the 3D fetal facial surface, 14 silicone target markers are pasted to the frontal facial surface of the fetal phantom, as shown in Fig. 7.8 (b), where all markers are the same hemispheres having a diameter of 5 mm and a height of 2.5 mm (Fig. 7.8 (a)). The 14 markers can easily be identified manually in 3D US images (Fig. 7.8 (c)). Mean TREs of these 14 markers for the test database are experimentally obtained as shown in Table 7.3, which proves good accuracy of the proposed method.

(a) One marker (b) 14 Markers (c) 3D US data

Figure 7.8 Fetal phantom with 14 markers and its US data.

Table 7.3 Mean TREs of each phantom (mm).

Fetus 1	Fetus 2	Fetus 3	Fetus 4	Fetus 5
1.52 ± 0.32	1.75 ± 0.43	1.46 ± 0.39	1.58 ± 0.64	1.44 ± 0.50
Mean TRE for all fetuses : 1.55 ± 0.46				

(2) Validation for Internal Organs

Although a high accuracy for locating the points on the 3D facial surface is obtained by measuring TREs, the location of internal organs, i.e. fetal oral cavity and airways, is still difficult. On the other hand, the geometric relationship between the internal organs and the 3D fetal facial surface in the 3D fetal model can be known based on their design. Therefore, the corresponding positions of the internal organs in real fetal phantoms can be estimated via such a geometric relationship.

Herein, as shown in Fig. 7.9 (a), position *P* on the bottom of the pharynx and position *T* on the entrance of the trachea are specified as validation examples, which are considered as the farthest points in the internal organs from the 3D fetal facial surface. The corresponding positions in the real fetal phantom are defined as *P′* and *T′*. Then, an validation method for the location errors of position *P* and *T* is proposed as follows:

Step-1 As shown in Fig. 7.9 (b), three positions, n_0, n_1 and n_2 are specified manually as reference points on the surface of the 3D fetal model, where n_0 is the position of the nose tip, and n_1 and n_2 are the positions of the two eyes' inner corners;

Step-2 Let v_i (i ∈ [1,2]) be the vectors from n_0 to n_i, and let v_p be the vector from n_0 to position *P*. Since the selected three positions and *P* position can be known from the design of the 3D fetal model, θ_i (i ∈ [1,2]), which are the angles between v_p and v_i, are computed. Meanwhile, the cross product $v_2 \times v_1$ is defined as a vector $v_{2\times1}$ that is perpendicular to both v_2 and v_1, with a direction given by the right-hand rule. Then, θ_p, which is the angle between v_p and $v_{2\times1}$ is computed;

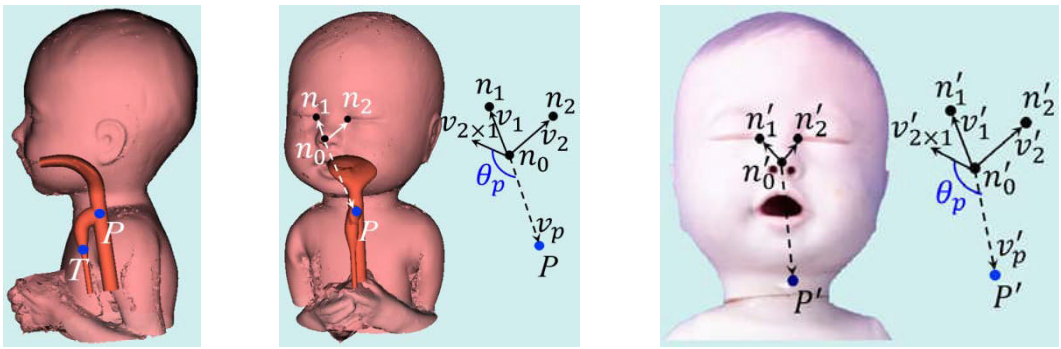
Step-3 The corresponding three positions n'_i (i ∈ [0,2]) on the surface of the real fetal phantom in 3D tracking space are directly measured by the pen probe prior to experiments. The vectors v'_i (i ∈ [1,2]) from n'_0 to n'_i and $v'_{2\times1}$, which is the cross product $v'_2 \times v'_1$ are computed, as shown in Fig. 7.9 (c). Since the geometric relationship among the selected three positions n_i (i ∈ [0,2]) and position *P* in the 3D fetal model is the same as that among the corresponding three position n'_i (i ∈ [0,2]) and position *P′* in the real fetal phantom, the following relationships hold.

$$v'_i \cdot v'_p = |v'_i| \cdot |v'_p| \cdot \cos\theta_i, \quad i \in [1,2] \tag{7.4}$$

$$v'_{2\times1} \cdot v'_p = |v'_{2\times1}| \cdot |v'_p| \cdot \cos\theta_p \tag{7.5}$$

$$\frac{|v'_1|}{|v'_p|} = \frac{|v_1|}{|v_p|} \tag{7.6}$$

where, θ_i (i ∈ [1,4]) and θ_p are calculated in Step-2, and the coordinate (x, y, z) of position P' in 3D tracking space are the solutions of the above simultaneous equations from Eq. (7.4) to Eq. (7.6). Thus, a unique solution of (x, y, z) is computed as a ground truth for estimating the position *P*.



(a) P & T (b) reference points (c) the corresponding points

Figure 7.9 The location errors of the pharynx and trachea.

Table 7.4 Location errors of the pharynx and trachea.

Positions	Location errors (mm)
Position *P* on pharynx	$2.51 \pm 0.47\ mm$
Position *T* on trachea	$3.04 \pm 0.59\ mm$

Table 7.5 The mean execution time of each key step (ms).

	Step-1	Step-2	Step-3
mean	15906 ± 341 mm	924 ± 96 mm	81 ± 32 mm
mean (the system)	16911 ± 469 mm		

On the other hand, the estimated coordinates of the position *P* from the 3D fetal model are transformed into the 3D tracking space based on the ICP-registration matrix

and 3D US calibration system. Then, the transformed coordinates in the 3D tracking space of the position P' are compared with the ground truth (the estimated coordinates of P) to calculate its location error. Similarly, the location error of position T can also be estimated. As a result of experiments, the mean location errors of the position P and T for the testing database are $2.51 \pm 0.47\ mm$ and $3.04 \pm 0.59\ mm$, respectively, as shown in Table 7.4.

In addition, for FETO surgeries, the surgical tool having an outer diameter of 2.4 mm is inserted into fetal airways until the tool reaches the entrance of fetal trachea. During this procedure, the narrowest part of the passage is the fetal pharynx, through which the surgical tool is supposed to go. The FETO surgery is normally performed at about 26-29 weeks gestational age (GA); therefore, the average diameter of fetal pharynx is 7.7 mm at that time. Thus, 3.85 mm - half of the average diameter of fetal pharynx is the maximum distance error tolerable for this surgery. Consequently, both of the location errors of the pharynx ($2.51 \pm 0.47\ mm$) and the entrance of the trachea ($3.04 \pm 0.59\ mm$) satisfy the requirement of the FETO surgery.

7.4.2 Efficiency of the Surgical Navigation System

In the online process, there are several factors contributing to the final efficiency of the surgical navigation system: (1) the extraction of 3D fetal facial surface from 3D US data (Step-1); (2) ICP-based registration, including key facial feature detection (Step-2); (3) 3D US calibration-based transformation (Step-3). The mean execution time of each key step for 10 groups of five fetal phantoms are listed in Table 7.5, which shows the system spends about $16911 \pm 469\ ms$ on the location of 3D fetal facial surface and internal organs. On one hand, the fetus is supposed to be immobilized during the FETO operation [23]. Once the registration is realized before the operation, no more registration is required during the operation so that real-time 3D registration is not necessary for the surgery. On the other hand, compared with 221 ± 69 minutes, which is medical doctors' average time for FETO surgeries [15], the time of the navigation system is very short and thus should meet current requirement of the FETO surgery.

However, the current survival rate in fetuses with left-sided CDH (the majority of hernias (80-85%) occur on the left side of the diaphragm) treated with the FETO surgery is not so high: the specific survival rate is between 24.1% and 49.1% [23]. Therefore, reducing the surgery time might be one solution to raise the survival rate.

Thus, a faster computation speed of the surgical navigation system could contribute to such a solution; in other words, faster computations are always demanded.

In addition, it can be learned that Step-1 spends the most time ($15906 \pm 341\ ms$) during the online process, which is a major factor deciding the final efficiency of the system. However, the efficiency of Step-1 can easily be increased by 50-100 times based on the technology of multi-threaded processing and multi-core processors. The efficiency of Step-2 and Step-3 can also be increased by 10-50 times based on GPU technology. Consequently, the computation time of the system can be improved to be less than one second in the future by making use of algorithm optimization, multi-threaded processing, multi-core processors, and GPU technology.

7.5 Conclusions

This chapter has proposed an approach that can accurately locate 3D fetal facial surface, oral cavity, and airways based on the registration between the 3D fetal model and 3D US images and a 3D US calibration system. The reconstructed 3D fetal model with the oral cavity and airways is registered with the 3D fetal facial surface extracted from 3D US images and located in 3D US space. Then, the 3D US calibration system based on a novel cones' phantom is proposed to transform the 3D fetal model from the 3D US space to the 3D tracking space, and a FRE of $1.49 \pm 0.44\ mm$ and a TRE of $1.81 \pm 0.56\ mm$ by using 24 fiducial points from two US volumes are achieved. Furthermore, a mean TRE of $1.55 \pm 0.46\ mm$ is achieved for measuring location accuracy of the 3D fetal facial surface extracted from 3D US images by 14 target markers, and mean location errors of $2.51 \pm 0.47\ mm$ and $3.04 \pm 0.59\ mm$ are respectively achieved for indirectly measuring location accuracy of the pharynx and the entrance of the trachea, which satisfies the requirement of the FETO surgery.

In the future, more points (e.g., some 3D points on the fetal body) will be introduced to reduce the error of ICP-based registration. Furthermore, the location of the surgical tool and the feedback from fetoscopic video sequences will be included in the surgical navigation system to make a comprehensive evaluation. Clinical fetal MR images are also essential for directly validating the location errors of the fetal oral cavity and airways, and will be considered in the cooperation with some hospitals.

Chapter 8 Conclusions and Future Work

8.1 Conclusions

To support FETO surgery's surgeons, this thesis aims at achieving a surgical navigation system, in which a flexible surgical tool for treating the fetus for his/her CDH is automatically or semi-automatically inserted into the fetal trachea. To facilitate the surgeons' judgments and confirmations, the navigation system should be able to provide information on spatial relationship between the surgical tool and fetal face, mouth and trachea for the surgeon in a robust, accurate and efficient manner. To satisfy this requirement, this thesis has proposed medical image processing methods that can detect the fetal face, reconstruct 3D fetal facial surface, locate the surgical tool, track the fetal mouth, and estimate the position of the fetal trachea. Experimental results prove that the proposed methods' achievements are within the tolerances determined by the FETO surgery in terms of robustness, accuracy and efficiency. In the following, specific conclusions for each method proposed in Chapters 5 to 7 are described.

(1) Fetal face detection

A method for detecting the fetal face from 3D US images has been proposed. A new boosting traversal scheme based on the mean and Gaussian curvatures of the surface detects five key facial features: nose tip, two eyes' inner corners, nose upper bridge and mouth. These detected key features are used for locating the fetal mouth from 3D US images and for performing coarse registration of 3D fetal facial surface. Experimental results show that all key feature points are successfully detected from a dataset of 72 3D US images for six fetal phantoms. The location error 3.18±0.91 mm of the detected upper lip is within the tolerance required by the FETO surgery. The detection system can complete the entire detection process within about 625 msec on a quad-core 2.60 GHz computer.

(2) Surgical tool location and fetal mouth tracking

To locate the flexible surgical tool, after 3D skeletons of the surgical tool are extracted from 3D US images, the coordinates of 15 points on the tool are obtained based on the 3D US calibration system (Section 4.3 , (c) in this section). Experimental

results show that the proposed approach can achieve sufficiently high accuracy for localizing the surgical tool: i.e., an average TRE of 1.99±0.49 mm and a maximum TRE of 2.61±0.49 mm, which satisfy the requirement of the FETO surgery: less than 3.85 mm.

To track the fetal mouth, first, the AdaBoost algorithm as a classifier detects the fetal face based on Haar-like features. Then, the CamShift algorithm is applied to track the fetal face based on the image color histogram, and the fetal mouth is detected as an ellipse by the improved IRHT method (Section 4.2 , (b) in this section). Experimental results confirm that the fetal mouth can be tracked accurately in real-time from fetoscopic video sequences.

(3) Estimating the position of the fetal trachea

The fetal phantom is registered to the 3D US images so that the internal organs including fetal trachea are located in the 3D tracking space based on the 3D US calibration system (Section 4.3 , (c) in this section). In experiments, a mean location error of 2.51±0.47 mm is achieved for localizing the fetal trachea in the 3D tracking space, which satisfies the requirement of the FETO surgery.

In addition, Chapter 4 explores the three basic technologies, which are used by the above-mentioned three technologies, as follows.

(a) Medical image segmentation

An improved Kernel-based fuzzy C-means algorithm is proposed to raise the segmentation accuracy for 2D US image segmentation by combining with spatial information. Experimental results indicate that the proposed method has a strong robustness for image noise, and also achieves a better segmentation performance than other fuzzy clustering methods. Also, it turns out that this segmentation method is useful as the pre-process for the fetal face detection (the above-mentioned (1)).

(b) Ellipse detection

An improved iterative randomized Hough transform (iIRHT) method is proposed to reduce the detection error and enhance the efficiency of the algorithm for detection of fetal head in 2D US images, fetal mouth in fetoscopic images, and position of each cone in a 3D US calibration system. Experiments on clinical fetal ultrasound images prove that the proposed iIRHT method achieves more robust and accurate results, and has a

better efficiency and performance than the RHT method and the original IRHT method.

(c) 3D Ultrasound Calibration System

A novel 3D US calibration system based on cones' phantom is proposed to promise a high location accuracy and robustness of the system for the location of the fetal facial surface extracted from 3D US data, the 3D fetal model with oral cavity and airways mapped with the extracted fetal facial surface, and the surgical tool detected from 3D US data. Experimental results are comparable to those obtained by previous calibration methods using EM tracking. The proposed system guarantees satisfying accuracy, and also only requires a simple experimental setup and two US volumes for 3D US calibration.

8.2 Future Work

There is a long way to realize a surgical navigation system (Fig. 1.8) for clinical applications or commercial use. As described in Section 8.1 , this thesis achieved some contributions to the prospective surgical navigation system, but each of the three proposed methods has remaining issues and future work as follows.

(1) Fetal face detection

Although a detection success rate of 100% is achieved by testing 72 images of 6 subjects, all of the test data are based on the fetal phantoms. Thus, the examination on the clinical data is more significant, and should be performed in the future.

(2) Surgical tool location and fetal mouth tracking

Robustness of the proposed method should be evaluated using more experimental data for surgical tool location. In addition, a method that can locate the tool even after the tool is inserted into the fetal mouth should be developed and explored.

Concerning the fetal mouth tracking from the fetoscopic images, dynamism of fetal mouth (open, close, etc.) has not been explored by this thesis. In particular, a method for dealing with the case in which the shape of the fetal mouth cannot be described by an ellipse needs to be developed. To solve this problem, multiple features' detection will be researched in the future for improving the robustness of the method on tracking fetal mouth.

(3) Estimating position of the fetal trachea

At present, the 3D fetal model with trachea is reconstructed based on the fetal phantoms and the designed structure of the trachea. In the future, the clinical 3D fetal MR or CT images will be utilized to reconstruct a 3D fetal model with trachea by collaborating with medical doctors and hospitals.

In addition, remaining issues and future work of the three basic technologies are described as follows.

(a) Medical image segmentation

Higher segmentation accuracies should be achieved for 2D US image segmentation; for that, the current algorithm should be improved based on studies using more US images. Also, the efficiency of the process should be enhanced in the future by algorithm optimization, because it needs to be used for all of 2D US slices (at least 256 slices) extracted from 3D US images.

(b) Ellipse detection

The accuracy and efficiency of the algorithm have been improved for detecting fetal head and fetal mouth in this thesis. In the future, more clinical ultrasound data (more than one hundred) will be tested for evaluating the accuracy, efficiency, and robustness of the proposed method.

(c) 3D US calibration

More 3D US volumes (more than thirty) will be collected to improve the precision and accuracy of the calibration system, which is the most important step for accuracy of the entire surgical navigation system.

Finally, considering the achievement of the final goal of the surgical navigation system, there are still some future works as follows.

First, in order to assess the technical feasibility and behavior of the system, the linkage of the surgical navigation system, the movement of the surgical tool and interaction with surgeons will be realized; thus, a feedback about location of the objects and interaction from the surgeons will be provided for guiding the movement of the surgical tool.

Second, although the experimental results show that the proposed methods attain

higher accuracy in average, a large number of cases have not yet been attempted for statistical significance test in this thesis. Thus, more cases of experiments will be performed, and more experimental data will be collected in the future for statistical significance test by comparison with previous methods to check the technical applicability and reliability of the system based on a clinical setting before clinical studies. Meanwhile, the system will be assessed for its clinical accuracy, patient and user safety, and reliability in a realistic clinical setting context.

Third, the system will be assessed in clinical trials, including study of the indirect and direct impacts on surgical strategy and performance by simulated experiments on fetal phantoms and animals.

Fourth, algorithm optimization, multi -threaded processing, multi-core processors, and GPU technology will be tried to heighten the efficiency of the proposed methods.

Last but not the least, clinical requirements of the current FETO surgery will further be confirmed to check whether the system is helpful for the surgery by discussion and collaboration with clinicians.

Acknowledgement

This dissertation summarizes the achievements of my research in Graduate School of Global Information and Telecommunication Studies, Waseda University, Tokyo, Japan, under the supervision of Prof. Jun Ohya. Firstly, I would like to express my grate acknowledgement to my supervisor, Prof. Ohya, for his valuable guidance, insightful suggestions, kind and constant help and encouragement during the past four years. I also would like to convey my sincere gratitude to my doctoral thesis review panel which included Prof. Wataru Kameyama, and Prof. Takashi Kawai, GITS, Waseda University, Tokyo, Japan, and Prof. Yoshinobu Sato, Osaka University, Osaka, Japan for their kind acceptance to be reviewers and their helpful comments.

I really appreciate Prof. Yoshinobu Sato in Osaka University, and Prof. Masakatsu G. Fujie in Waseda University for their kind help, many valuable comments and meaningful discussions. I also really thank Dr. Bo Zhang, and Mr. Quanquan Liu, in Fujie Lab, and Mr. Chunbao Wang in Takanishi Lab, Faculty of Science and Engineering, Waseda University, for their great supports in experiments.

I would like to thank Dr. Toshio Chiba, deputy director of the National Center for Child Health and Development, and Dr. Hiromasa Yamashita, clinical fellow of the National Center for Child Health and Development, Tokyo, Japan, for their friendly help and valuable advice on clinical surgeries.

I want to thank my supervisor in the master course, Prof. Limin Luo, in Southeast University, Nanjing, China, for his continues encouragement and helps. In addition, I would like to give my thanks to all the members in Ohya Lab, Waseda University, especially, Dr. Yingdi Xie, Ms. Dan Luo, Mr. Tianliang Xie, Ms. Mengyao Duan, Ms. Ye Li, for their continuous encouragement, support, and technical discussions. Finally, I would like to thank all my friends in GITS, Waseda University.

I am very proud of being awarded the CSC scholarship from China Scholarship Council. I would like to thank for the financial support.

Last but not least, I would like to show my respects and deep gratitude to my

parents, my wife, my daughter and other family members. Their understanding, encouragement and support, have helped me to complete this work.

Rong Xu

April 2014

Reference

[1] I.M. Germano, "The NeuroStation System for image-guided, frameless stereotaxy," *Neurosurgery,* vol. 37, pp. 348-350, 1995.

[2] W.E.L. Grimson*, et al.*, "An automatic registration method for frameless stereotaxy, image guided surgery, and enhanced reality visualization," *IEEE Transactions on Medical Imaging,* vol. 15, pp. 129-140, 1996.

[3] R. Kikinis*, et al.*, "Computer-assisted interactive three-dimensional planning for neurosurgical procedures," *Neurosurgery,* vol. 38, pp. 640-651, 1996.

[4] A. DiGioia, *Computer and Robotic Assisted Hip and Knee Surgery*: Oxford university press, 2004.

[5] L.P. Nolte and R. Ganz, *Computer-Assisted Orthopedic Surgery (CAOS)*: Hogrefe & Huber Publishers, 1999.

[6] Z. Yaniv and K. Cleary, "Image-guided procedures: A review," *Computer Aided Interventions and Medical Robotics,* vol. 3, pp. 1-63, 2006.

[7] R.A. Robb*, et al.*, "Computer-aided surgery planning and rehearsal at Mayo Clinic," *IEEE Computer,* vol. 29, pp. 39-47, 1996.

[8] A. Gaxiola*, et al.*, "Congenital diaphragmatic hernia: an overview of the etiology and current management," *Acta Pædiatrica,* vol. 98, pp. 621-627, 2009.

[9] M.R. Harrison*, et al.*, "A randomized trial of fetal endoscopic tracheal occlusion for severe fetal congenital diaphragmatic hernia," *New England Journal of Medicine,* vol. 349, pp. 1916-1924, 2003.

[10] E. Garne*, et al.*, "Congenital diaphragmatic hernia: evaluation of prenatal diagnosis in 20 European regions," *Ultrasound in Obstetrics & Gynecology,* vol. 19, pp. 329-333, 2002.

[11] J. Deprest*, et al.*, "Technical aspects of fetal endoscopic tracheal occlusion for congenital diaphragmatic hernia," *Journal of pediatric surgery,* vol. 46, pp. 22-32, 2011.

[12] M. Harrison*, et al.*, "Correction of congenital diaphragmatic hernia in utero. II. Simulated correction permits fetal lung growth with survival at birth," *Surgery,* vol. 88, pp. 260-268, 1980.

[13] P.A. Khan*, et al.*, "Tracheal occlusion: a review of obstructing fetal lungs to make them grow and mature," *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, vol. 145, pp. 125-138, 2007.

[14] A.W. Flake*, et al.*, "Treatment of severe congenital diaphragmatic hernia by fetal tracheal occlusion: clinical experience with fifteen cases," *American journal of obstetrics and gynecology,* vol. 183, pp. 1059-1066, 2000.

[15] M.R. Harrison*, et al.*, "Fetoscopic temporary tracheal occlusion for congenital diaphragmatic hernia: prelude to a randomized, controlled trial," *Journal of pediatric surgery,* vol. 38, pp. 1012-1020, 2003.

[16] F.I. Luks*, et al.*, "Fetoscopy-guided fetal endoscopy in a sheep model," *Journal of the American College of Surgeons,* vol. 178, pp. 609, 1994.

[17] J.A.M. Deprest*, et al.*, "Tracheoscopic endoluminal plugging using an inflatable device in the fetal lamb model," *European Journal of Obstetrics & Gynecology and Reproductive Biology,* vol. 81, pp. 165-169, 1998.

[18] M.R. Harrison, *et al.*, "Correction of congenital diaphragmatic hernia in utero: VI. Hard-earned lessons," *Journal of pediatric surgery,* vol. 28, pp. 1411-1418, 1993.

[19] A. David, *et al.*, "Ultrasound‐guided injection and occlusion of the trachea in fetal sheep," *Ultrasound in Obstetrics & Gynecology,* vol. 28, pp. 82-88, 2006.

[20] M.R. Harrison, *et al.*, "Fetoscopic temporary tracheal occlusion by means of detachable balloon for congenital diaphragmatic hernia," *American journal of obstetrics and gynecology,* vol. 185, pp. 730-733, 2001.

[21] R. Quintero, *et al.*, "ORIGINAL ARTICLES-Minimally invasive intraluminal tracheal occlusion in a human fetus with left congenital diaphragmatic hernia at 27 weeks' gestation via direct fetal laryngoscopy," *Prenatal and Neonatal Medicine,* vol. 5, pp. 134-140, 2000.

[22] J. Deprest, *et al.*, "Fetoscopic tracheal occlusion (FETO) for severe congenital diaphragmatic hernia: evolution of a technique and preliminary results," *Ultrasound in Obstetrics & Gynecology,* vol. 24, pp. 121-126, 2004.

[23] J. Jani, *et al.*, "Severe diaphragmatic hernia treated by fetal endoscopic tracheal occlusion," *Ultrasound in Obstetrics & Gynecology,* vol. 34, pp. 304-310, 2009.

[24] H. Flageole, *et al.*, "The plug-unplug sequence: an important step to achieve type II pneumocyte maturation in the fetal lamb model," *Journal of pediatric surgery,* vol. 33, pp. 299-303, 1998.

[25] B. Zhang, *et al.*, "Development of 6-DOF wire-driven robotic manipulator for minimally invasive fetal surgery," *IEEE International Conference on Robotics and Automation (ICRA)*, Shanghai, pp. 2892-2897, 2011.

[26] B. Zhang, *et al.*, "Development of a robotic manipulator system for Congenital Diaphragmatic Hernia," *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Anchorage, AK, pp. 723-728, 2011.

[27] T.M. Peters, "Image-guided surgery: from X-rays to virtual reality," *Computer methods in biomechanics and biomedical engineering,* vol. 4, pp. 27-57, 2001.

[28] R.L. Galloway Jr, "The process and development of image-guided procedures," *Annual review of biomedical engineering,* vol. 3, pp. 83-108, 2001.

[29] I. Bankman, *Handbook of medical image processing and analysis*: Access Online via Elsevier, 2008.

[30] J.D. Foley, *Computer graphics: Principles and practice in C* vol. 12110: Addison-Wesley Professional, 1996.

[31] D.P. Luebke, "A developer's survey of polygonal simplification algorithms," *IEEE Computer Graphics and Applications,* vol. 21, pp. 24-35, 2001.

[32] R. Ruano, *et al.*, "Four-dimensional ultrasonographic guidance of fetal tracheal occlusion in a congenital diaphragmatic hernia," *Journal of ultrasound in medicine,* vol. 26, pp. 105-109, 2007.

[33] M. Tchirikov, "Successful tracheal occlusion using ultrathin fetoscopic equipment combined with real-time three-dimensional ultrasound," *European Surgical Research,* vol. 43, pp. 204-207, 2009.

[34] G.D. Stetten and V.S. Chib, "Overlaying ultrasonographic images on direct vision," *Journal of ultrasound in medicine,* vol. 20, pp. 235-240, 2001.

[35] G. Fichtinger, *et al.*, "Image overlay guidance for needle insertion in CT scanner," *IEEE Transactions on Biomedical Engineering,* vol. 52, pp. 1415-1424, 2005.

[36] G.S. Fischer, *et al.*, "MRI image overlay: applications to arthrography needle insertion," *Studies in*

health technology and informatics, vol. 119, pp. 150-155, 2005.

[37] K.T. Foley*, et al.*, "Virtual fluoroscopy: Multi-planar x-ray guidance with minimal radiation exposure," *European Spine Journal,* vol. 8, pp. 36, 1999.

[38] R. Hofstetter*, et al.*, "Fluoroscopy as an imaging means for computer-assisted surgical navigation," *Computer aided surgery,* vol. 4, pp. 65-76, 1999.

[39] V. Rohde*, et al.*, "Advanced neuronavigation in skull base tumors and vascular lesions," *min-Minimally Invasive Neurosurgery,* vol. 48, pp. 13-18, 2005.

[40] D.A. Feinberg and A.S. Mark, "Human brain motion and cerebrospinal fluid circulation demonstrated with MR velocity imaging," *Radiology,* vol. 163, pp. 793-799, 1987.

[41] C. de Graaf and M. Viergever, *Information processing in medical imaging*: Plenum Press, 1987.

[42] A. Rosenfeld and A.C. Kak, *Digital picture processing* vol. 1: Elsevier, 1982.

[43] I.W. Selesnick and C.S. Burrus, "Generalized digital Butterworth filter design," *IEEE Transactions on Signal Processing,* vol. 46, pp. 1688-1694, 1998.

[44] F. Jin*, et al.*, "Adaptive Wiener filtering of noisy images and image sequences," *Proceedings of International Conference on Image Processing 2003*, vol. 3, pp. III-349-52 vol. 2, 2003.

[45] D.L. Pham*, et al.*, "Current methods in medical image segmentation," *Annual review of biomedical engineering,* vol. 2, pp. 315-337, 2000.

[46] Nicholas Ayache*, et al.*, "Segmentation of complex three-dimensional medical objects: A challange and a requirement for computer-assisted surgery planning and performance," in *Computer-integrated surgery, Technology and clinical applications*, S.L. R. H. Taylor, G. C. Burdea, and R. MÄosges, Ed., MIT Press, Cambridge Ma. ed, 1995, pp. 59-74.

[47] J.S. Duncan and N. Ayache, "Medical image analysis: Progress over two decades and the challenges ahead," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, pp. 85-106, 2000.

[48] T. McInerney and D. Terzopoulos, "Deformable models in medical image analysis: a survey," *Medical image analysis,* vol. 1, pp. 91-108, 1996.

[49] J. Montagnat*, et al.*, "A review of deformable surfaces: topology, geometry and deformation," *Image and Vision Computing,* vol. 19, pp. 1023-1040, 2001.

[50] J.S. Suri*, et al.*, "Shape recovery algorithms using level sets in 2-D/3-D medical imagery: a state-of-the-art review," *IEEE Transactions on Information Technology in Biomedicine,* vol. 6, pp. 8-28, 2002.

[51] C. Kirbas and F. Quek, "A review of vessel extraction techniques and algorithms," *ACM Computing Surveys (CSUR),* vol. 36, pp. 81-121, 2004.

[52] J.S. Suri*, et al.*, "A review on MR vascular image processing algorithms: acquisition and prefiltering: part I," *IEEE transactions on information technology in biomedicine,* vol. 6, pp. 324-337, 2002.

[53] J.S. Suri*, et al.*, "A review on MR vascular image processing: skeleton versus nonskeleton approaches: part II," *IEEE transactions on information technology in biomedicine,* vol. 6, pp. 338-350, 2002.

[54] J.S. Suri*, et al.*, "Computer vision and pattern recognition techniques for 2-D and 3-D MR cerebral cortical segmentation (part I): A state-of-the-art review," *Pattern Analysis & Applications,* vol. 5, pp. 46-76, 2002.

[55] J.S. Suri*, et al.*, "Fusion of region and boundary/surface-based computer vision and pattern

recognition techniques for 2-D and 3-D MR cerebral cortical segmentation (Part-II): A state-of-the-art review," *Pattern Analysis & Applications,* vol. 5, pp. 77-98, 2002.

[56] R. Xu*, et al.*, "Segmentation of Brain MRI," in *Advances in Brain Imaging*, V. Chaudhary, Ed., ed: InTech, 2012, pp. 1-29.

[57] M. Descoteaux*, et al.*, "Bone enhancement filtering: Application to sinus bone segmentation and simulation of pituitary surgery," *Proceedings of Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, pp. 9-16, 2005.

[58] M. Kass*, et al.*, "Snakes: Active contour models," *International journal of computer vision,* vol. 1, pp. 321-331, 1988.

[59] P. Yan and A.A. Kassim, "MRA image segmentation with capillary active contour," *Proceedings of Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, pp. 51-58, 2005.

[60] J.C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*: Plenum Press, 1981.

[61] D.Q. Zhang and S.C. Chen, "Kernel-based fuzzy and possibilistic c-means clustering," *Proceedings of the International Conference Artificial Neural Network*, pp. 122-125, 2003.

[62] S. Kannan*, et al.*, "New Robust Fuzzy C-Means Based Gaussian Function in Classifying Brain Tissue Regions," in *Contemporary Computing*, ed: Springer, 2009, pp. 158-169.

[63] A.W.-C. Liew and H. Yan, "An adaptive spatial fuzzy clustering algorithm for 3-D MR image segmentation," *IEEE Transactions on Medical Imaging,* vol. 22, pp. 1063-1075, 2003.

[64] D.-Q. Zhang and S.-C. Chen, "A novel kernelized fuzzy c-means algorithm with application in medical image segmentation," *Artificial Intelligence in Medicine,* vol. 32, pp. 37-50, 2004.

[65] K.-S. Chuang*, et al.*, "Fuzzy c-means clustering with spatial information for image segmentation," *Computerized medical imaging and graphics,* vol. 30, pp. 9-15, 2006.

[66] X.F. Yu*, et al.*, "A New Algorithm for Magnetic Resonance Image Segmentation based on Fuzzy Kernel Clustering," *Journal of Southern Medical University,* vol. 28, pp. 555-557, 2008.

[67] M.-S. Yang and H.-S. Tsai, "A Gaussian kernel-based fuzzy C-means algorithm with a spatial bias correction," *Pattern Recognition Letters,* vol. 29, pp. 1713-1725, 2008.

[68] S. Ramathilagam*, et al.*, "Modified fuzzy c-means algorithm for segmentation of T1–T2-weighted brain MRI," *Journal of Computational and Applied Mathematics,* vol. 235, pp. 1578-1586, 2011.

[69] R. Xu and J. Ohya, "An improved Kernel-based Fuzzy C-means Algorithm with spatial information for brain MR image segmentation," *Proceedings of the 25th International Conference of Image and Vision Computing New Zealand (IVCNZ)*, pp. 1-7, 2010.

[70] A. Fitzgibbon*, et al.*, "Direct least square fitting of ellipses," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 21, pp. 476-480, 1999.

[71] P.L. Rosin, "A note on the least squares fitting of ellipses," *Pattern Recognition Letters,* vol. 14, pp. 799-808, 1993.

[72] H.K. Yuen*, et al.*, "Detecting Partially Occluded Ellipses Using The Hough Transform," *Image and Vision Computing,* vol. 7, pp. 31-37, Feb 1989.

[73] D. Ben-Tzvi and M. Sandler, "A combinatorial Hough transform," *Pattern Recognition Letters,* vol. 11, pp. 167-174, 1990.

[74] L. Xu and E. Oja, "Randomized Hough transform (RHT): basic mechanisms, algorithms, and computational complexities," *CVGIP Image Understanding,* vol. 57, pp. 131-131, 1993.

[75] N. Kiryati, *et al.*, "A probabilistic Hough transform," *Pattern Recognition,* vol. 24, pp. 303-316, 1991.

[76] V.F. Leavers, "The dynamic generalized Hough transform: its relationship to the probabilistic Hough transforms and an application to the concurrent detection of circles and ellipses," *CVGIP: Image understanding,* vol. 56, pp. 381-398, 1992.

[77] M.A. Fischler and R.C. Bolles, "Random Sample Consensus - A Paradigm For Model Fitting With Applications To Image Analysis And Automated Cartography," *Communications of the Acm,* vol. 24, pp. 381-395, 1981.

[78] G. Gordon, "Face recognition based on depth maps and surface curvature," *Proc. of SPIE Conf. on Geometric Methods in Computer Vision*, vol. 1570, pp. 234-247, 1991.

[79] H.T. Tanaka, *et al.*, "Curvature-based face surface recognition using spherical correlation. principal directions for curved object recognition," *Proc. FG' 98*, pp. 372-377, 1998.

[80] A.B. Moreno, *et al.*, "Face recognition using 3D surface-extracted descriptors," *Proc. of the Irish Machine Vision and Image Processing*, vol. 2003, pp. 1-8, 2003.

[81] A. Colombo, *et al.*, "3D face detection using curvature analysis," *Pattern Recognition,* vol. 39, pp. 444-455, 2006.

[82] B. Amberg, *et al.*, "Expression invariant 3D face recognition with a morphable model," *Proc. FG' 08*, pp. 1-6, 2008.

[83] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. on PAMI,* vol. 25, pp. 1063-1074, 2003.

[84] C. Hesher, *et al.*, "A novel technique for face recognition using range imaging," *7th International Symposium on Signal Processing and Its Applications*, vol. 2, pp. 201-204, 2003.

[85] X. Lu, *et al.*, "Matching 2.5 D face scans to 3D models," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, pp. 31-43, 2006.

[86] C.S. Chua, *et al.*, "3D human face recognition using point signature," *Proc. FG' 00*, pp. 233-238, 2000.

[87] C. Conde and Á. Serrano, "3D facial normalization with spin images and influence of range data calculation over face verification," *Proc. CVPR' 05*, pp. 115-200, 2005.

[88] C. Beumier and M. Acheroy, "Automatic 3D face authentication," *Image and Vision Computing,* vol. 18, pp. 315-321, 2000.

[89] C. Samir, *et al.*, "Three-dimensional face recognition using shapes of facial curves," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, pp. 1858-1863, 2006.

[90] S. Feng, *et al.*, "Automatic fetal face detection from ultrasound volumes via learning 3D and 2D information," *Proc. CVPR' 09*, pp. 2488-2495, 2009.

[91] J.M. Fitzpatrick, *et al.*, "Predicting error in rigid-body point-based registration," *IEEE Transactions on Medical Imaging,* vol. 17, pp. 694-702, 1998.

[92] P.M. Thompson, *et al.*, "Growth patterns in the developing brain detected by using continuum mechanical tensor maps," *Nature,* vol. 404, pp. 190-193, 2000.

[93] H. Jiang, *et al.*, "New approach to 3-D registration of multimodality medical images by surface matching," *Visualization in biomedical computing*, pp. 196-213, 1992.

[94] P.J. Besl and N.D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern*

Analysis and Machine Intelligence, vol. 14, pp. 239-256, 1992.

[95] P.A. Van den Elsen*, et al.*, "Automatic registration of CT and MR brain images using correlation of geometrical features," *IEEE Transactions on Medical Imaging,* vol. 14, pp. 384-396, 1995.

[96] J.V. Hajnal*, et al.*, "A registration and interpolation procedure for subvoxel matching of serially acquired MR images," *Journal of Computer Assisted Tomography,* vol. 19, pp. 289-296, 1995.

[97] A. Kassam and M.L. Wood, "Fourier registration of three‐dimensional brain MR images: Exploiting the axis of rotation," *Journal of Magnetic Resonance Imaging,* vol. 6, pp. 894-902, 1996.

[98] R.P. Woods*, et al.*, "Automated image registration: I. General methods and intrasubject, intramodality validation," *Journal of Computer Assisted Tomography,* vol. 22, pp. 139-152, 1998.

[99] R.P. Woods*, et al.*, "MRI-PET registration with automated algorithm," *Journal of Computer Assisted Tomography,* vol. 17, pp. 536-546, 1993.

[100] C.E. Shannon, "The mathematical theory of communication (parts 1 and 2)," *Bell System Technical Journal,* vol. 27, pp. 379-423 and 623-656, 1949.

[101] F. Maes*, et al.*, "Multimodality image registration by maximization of mutual information," *IEEE Transactions on Medical Imaging,* vol. 16, pp. 187-198, 1997.

[102] C. Yang and G. Medioni, "Object modelling by registration of multiple range images," *Image and Vision Computing,* vol. 10, pp. 145-155, 1992.

[103] H. Visarius*, et al.*, "Man-machine interfaces in computer assisted surgery," *Computer aided surgery,* vol. 2, pp. 102-107, 1997.

[104] W.M. Wells III*, et al.*, "Adaptive segmentation of MRI data," *IEEE Transactions on Medical Imaging,* vol. 15, pp. 429-442, 1996.

[105] B. Johnston*, et al.*, "Segmentation of multiple sclerosis lesions in intensity corrected multispectral MRI," *IEEE Transactions on Medical Imaging,* vol. 15, pp. 154-169, 1996.

[106] D.L. Pham and J.L. Prince, "An adaptive fuzzy C-means algorithm for image segmentation in the presence of intensity inhomogeneities," *Pattern Recognition Letters,* vol. 20, pp. 57-68, 1999.

[107] J.K. Udupa and S. Samarasekera, "Fuzzy connectedness and object definition: theory, algorithms, and applications in image segmentation," *Graphical Models and Image Processing,* vol. 58, pp. 246-261, 1996.

[108] G.C. Karmakar and L.S. Dooley, "A generic fuzzy rule based image segmentation algorithm," *Pattern Recognition Letters,* vol. 23, pp. 1215-1227, 2002.

[109] F. Klauschen*, et al.*, "Evaluation of automated brain MR image segmentation and volumetry methods," *Human Brain Mapping,* vol. 30, pp. 1310-1327, 2009.

[110] F.P. Hadlock, "Ultrasound evaluation of fetal growth," *In: Callen P, ed. Ultrasonography in obstetrics and gynecology, Philadelphia, PA: Saunders,* pp. 129-142, 1988.

[111] I. Gull*, et al.*, "Prediction of fetal weight by ultrasound: the contribution of additional examiners," *Ultrasound in Obstetrics and Gynecology,* vol. 20, pp. 57-60, 2002.

[112] R. Xu*, et al.*, "Automatic real-time tracking of fetal mouth in fetoscopic video sequence for supporting fetal surgeries," *SPIE Medical Imaging*, pp. 86710Z-86710Z-6, 2013.

[113] J. Illingworth and J. Kittler, "A survey of the Hough transform," *Computer vision, graphics, and image processing,* vol. 44, pp. 87-116, 1988.

[114] W.-Y. Wu and M.-J.J. Wang, "Elliptical object detection by using its geometric properties," *Pattern*

Recognition, vol. 26, pp. 1499-1509, 1993.

[115] W. Lu and J.L. Tan, "Detection of ellipse in images with strong noise by iterative randomized Hough transform (IRHT)," *Pattern Recognition,* vol. 41, pp. 1268-1279, Apr 2008.

[116] W. Lu*, et al.*, "Automated fetal head detection and measurement in ultrasound images by iterative randomized Hough transform," *Ultrasound in Medicine and Biology,* vol. 31, pp. 929-936, Jul 2005.

[117] L.-f. LI*, et al.*, "A randomized algorithm for detecting multiple ellipses based on least square approach," *Opto-electronics Review,* vol. 13, pp. 61-67, 2005.

[118] R. Xu*, et al.*, "Automatic Fetal Head Detection on Ultrasound Images by An Improved Iterative Randomized Hough Transform," *Proc. of IVCNZ' 11*, Auckland, pp. 288-292, 2011.

[119] R. Xu*, et al.*, "Quasi Random Scheme Based Improved Iterative Randomized Hough Transform (IRHT) for Detection of Partial Ellipses in Medical Images," *IIEEJ Transactions on Image Electronics and Visual Computing,* vol. 1, pp. 97-106, 2013.

[120] S. Chang, "Extracting Skeletons from Distance Maps," *International Journal of Computer Science and Network Security,* vol. 7, pp. 213-219, 2007.

[121] K. Hahn*, et al.*, "Ellipse detection using a randomized Hough transform based on edge segment merging scheme," *Proceedings of the 6th WSEAS International Conference on Signal Processing, Robotics and Automation*, pp. 1-6, 2007.

[122] D.K. Prasad*, et al.*, "Edge curvature and convexity based ellipse detection method," *Pattern Recognition,* vol. 45, pp. 3204-3221, 2012.

[123] D.K. Prasad*, et al.*, "An error bounded tangent estimator for digitized elliptic curves," *Discrete Geometry for Computer Imagery*, pp. 272-283, 2011.

[124] L. Xu*, et al.*, "A New Curve Detection Method - Randomized Hough Transform (RHT)," *Pattern Recognition Letters,* vol. 11, pp. 331-338, May 1990.

[125] A. Forbes, "Fitting an ellipse to data," *National Physical Laboratory Report DITC 95/87,* 1987.

[126] F.P. Hadlock*, et al.*, "Estimating fetal age: effect of head shape on BPD," *American Journal of Roentgenology,* vol. 137, pp. 83-85, 1981.

[127] T.C. Poon and R.N. Rohling, "Comparison of calibration methods for spatial tracking of a 3-D ultrasound probe," *Ultrasound in medicine & biology,* vol. 31, pp. 1095-1108, 2005.

[128] C. Bergmeir*, et al.*, "Comparing calibration approaches for 3D ultrasound probes," *International journal of computer assisted radiology and surgery,* vol. 4, pp. 203-213, 2009.

[129] A. Lang*, et al.*, "Calibration of 3D ultrasound to an electromagnetic tracking system," *Proceedings of SPIE*, vol. 7968, pp. 79680W, 2011.

[130] H. Zhang*, et al.*, "Freehand 3D ultrasound calibration using an electromagnetically tracked needle," *Proceedings of SPIE*, vol. 6141, pp. 61412M, 2006.

[131] A. Khamene and F. Sauer, "A novel phantom-less spatial and temporal ultrasound calibration method," *Proceedings of MICCAI 2005,* vol. 3750, pp. 65-72, 2005.

[132] L.G. Bouchet*, et al.*, "Calibration of three-dimensional ultrasound images for image-guided radiation therapy," *Physics in Medicine and Biology,* vol. 46, pp. 559-577, 2001.

[133] Michael B.*, et al.*, "3D Ultrasound probe calibration for computer-guided diagnosis and therapy," *Dans Proceedings of CVAMIA'06,* pp. arXiv:0801.3711v1, 2008.

[134] T. Lange and S. Eulenstein, "Calibration of swept-volume 3-D ultrasound," *Medical Image*

Understanding and Analysis (MIUA), vol. 99, pp. 29-32, 2002.

[135] X. Huang*, et al.*, "Image registration based 3D TEE-EM calibration," *ISBI 2010*, pp. 1209-1212, 2010.

[136] V. Parthasarathy*, et al.*, "Real-time 3D ultrasound guided interventional system for cardiac stem cell therapy with motion Compensation," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011,* vol. 14, pp. 283-290, 2011.

[137] E.M. Boctor*, et al.*, "Ultrasound self-calibration," *Proceeds of SPIE*, vol. 6141, pp. 61412N, 2006.

[138] R. Prager*, et al.*, "Rapid calibration for 3-D freehand ultrasound," *Ultrasound in medicine & biology,* vol. 24, pp. 855-869, 1998.

[139] J. Blackall*, et al.*, "An image registration approach to automated calibration for freehand 3D ultrasound," *Proceedings of MICCAI 2000*, vol. 1935/2000, pp. 462-471, 2000.

[140] P.W. Hsu*, et al.*, "Freehand 3D ultrasound calibration: a review," *Advanced Imaging in Biology and Medicine,* vol. 1, pp. 47-84, 2008.

[141] P.W. Hsu*, et al.*, "Comparison of freehand 3-D ultrasound calibration techniques using a stylus," *Ultrasound in medicine & biology,* vol. 34, pp. 1610-1621, 2008.

[142] B.K.P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America A,* vol. 4, pp. 629-642, 1987.

[143] K.W. Bowyer*, et al.*, "A survey of approaches and challenges in 3D and multi-modal 3D+ 2D face recognition," *Computer Vision and Image Understanding,* vol. 101, pp. 1-15, 2006.

[144] A.F. Abate*, et al.*, "2D and 3D face recognition: A survey," *Pattern Recognition Letters,* vol. 28, pp. 1885-1906, 2007.

[145] D. Smeets*, et al.*, "Objective 3D face recognition: Evolution, approaches and challenges," *Forensic science international,* vol. 201, pp. 125-132, 2010.

[146] R. Xu*, et al.*, "Automatic Fetal Face Detection By Locating Fetal Facial Features From 3D Ultrasound Images For Navigating Fetoscopic Tracheal Occlusion Surgeries," *Proceedings of 10th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2013)*, Shanghai, China, pp. 1-6, 2013.

[147] P.J. Besl and R.C. Jain, "Invariant surface characteristics for 3D object recognition in range images," *Computer vision, graphics, and image processing,* vol. 33, pp. 33-80, 1986.

[148] M.P. Do Carmo, *Differential geometry of curves and surfaces*: Prentice-Hall Englewood Cliffs, NJ, 1976.

[149] G.G. Gordon, "Face recognition based on depth and curvature features," *Proc. CVPR' 92*, pp. 808-810, 1992.

[150] C. Xu*, et al.*, "Combining local features for robust nose location in 3D facial data," *Pattern Recognition Letters,* vol. 27, pp. 1487-1494, 2006.

[151] X. Lu and A.K. Jain, "Multimodal facial feature extraction for automatic 3D face recognition," *Tech. Rep. MSU-CSE-05-22, Michigan State University,* 2005.

[152] J.J. Koenderink and A.J. van Doorn, "Surface shape and curvature scales," *Image and Vision Computing,* vol. 10, pp. 557-564, 1992.

[153] D. Mueller, "Fast Marching Minimal Path Extraction in ITK," *The Insight Journal,* pp. 1-9, 2008.

[154] S. Tez*, et al.*, "Ultrasound evaluation of the normal fetal pharynx," *Early human development,* vol.

81, pp. 629-633, 2005.

[155] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. I-511-I-518, 2001.

[156] J.G. Allen*, et al.*, "Object tracking using camshift algorithm and multiple quantized feature spaces," *Proceedings of the Pan-Sydney area workshop on Visual information processing*, pp. 3-7, 2004.

[157] Y. Freund and R.E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," *Computational learning theory*, pp. 23-37, 1995.

[158] D. Levine, *Atlas of fetal MRI*: Taylor & Francis, 2005.

[159] J.L. Miller*, et al.*, "Emergence of oropharyngeal, laryngeal and swallowing activity in the developing fetal upper aerodigestive tract: an ultrasound evaluation," *Early human development,* vol. 71, pp. 61-87, 2003.

[160] K.D. Kalache*, et al.*, "Ultrasound measurements of the diameter of the fetal trachea, larynx and pharynx throughout gestation and applicability to prenatal diagnosis of obstructive anomalies of the upper respiratory–digestive tract," *Prenatal diagnosis,* vol. 19, pp. 211-218, 1999.

[161] G. Žauhar*, et al.*, "Comparison of the acoustic streaming in amniotic fluid and water in medical ultrasonic beams," *Ultraschall in der Medizin-European Journal of Ultrasound,* vol. 28, pp. 152-158, 2006.

[162] C. Kuhn*, et al.*, "Impact of extracorporeal shock waves on the human skin with cellulite: A case study of an unique instance," *Clinical interventions in aging,* vol. 3, pp. 201-210, 2008.

[163] W.R. Hendee and E.R. Ritenour, *Medical imaging physics*: John Wiley & Sons, 2003.

[164] G. Tonni*, et al.*, "Prenatal screening for fetal face and clefting in a prospective study on low-risk population: can 3-and 4-dimensional ultrasound enhance visualization and detection rate?," *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology,* vol. 100, pp. 420-426, 2005.

[165] S. Du*, et al.*, "An extension of the ICP algorithm considering scale factor," *IEEE International Conference on Image Processing (ICIP)*, vol. 5, pp. 193-196, 2007.

[166] C. Dorai and A.K. Jain, "COSMOS-A representation scheme for 3D free-form objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 19, pp. 1115-1130, 1997.

[167] A.N. Ansari*, et al.*, "3D Face mesh modeling from range images for 3D face recognition," *Proc. ICIP' 07*, vol. 4, pp. IV-509-IV-512, 2007.

Publication List

Category	Publication List
Journal Papers	○***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "Quasi Random Scheme Based Improved Iterative Randomized Hough Transform (IRHT) for Detection of Partial Ellipses in Medical Images", IIEEJ Transactions on Image Electronics and Visual Computing, vol. 1, no. 1, pp. 97-106, Dec., 2013. (Relate to Chapter 4)
	○***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "Locating Fetal Facial Surface, Oral Cavity and Airways by a 3D Ultrasound Calibration Using A Novel Cones' Phantom", IEICE Transactions on Information and Systems, vol.E97-D, no.5, pp.-, May, 2014. (Relate to Chapter 7)
International Conferences (Reviewed)	○***Rong Xu*** and Jun Ohya, "An Improved Kernel-based Fuzzy C-means Algorithm with Spatial Information for Brain MR Image Segmentation", Proceedings of the 25th International Conference of Image and Vision Computing New Zealand (IVCNZ' 2010), Queenstown, No.173, pp.1-7, 2010. (Relate to Chapter 4)
	○***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "Automatic Fetal Head Detection on Ultrasound Images by An Improved Iterative Randomized Hough Transform", Proceedings of the 26th International Conference of Image and Vision Computing New Zealand (IVCNZ' 2011), Auckland, pp.288-292, ISBN: 978-0-473-20281-1, 2011. (Relate to Chapter 4)
	○***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "A Flexible Surgical Tool Localization Using A 3D Ultrasound Calibration System for Fetoscopic Tracheal Occlusion (FETO)", Proceedings of MICCAI Workshop on Clinical Image-Based Procedures: From Planning to Intervention (CLIP 2012), Nice, France, Lecture Notes in Computer Science, vol.7761, pp.17-24, 2012. (Relate to Chapter 6)
	○***Rong Xu***, Tianliang Xie, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "Automatic real-time tracking of fetal mouth in fetoscopic video sequence for supporting fetal surgeries", Proceedings of SPIE Medical Imaging 2013: Image-Guided Procedures, Robotic Interventions, and Modeling, Orlando, FL, USA, vol.8671, 86710Z-86710Z-6, 2013. (Relate to Chapter 6)

	○***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "Automatic Fetal Face Detection By Locating Fetal Facial Features From 3D Ultrasound Images For Navigating Fetoscopic Tracheal Occlusion Surgeries", Proceedings of the 10th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2013), Shanghai, China, pp.1-6, 2013. (Relate to Chapter 5)
International Conferences (Reviewed, Co-author)	[1] Mengyao Duan, ***Rong Xu***, and Jun Ohya, "3D surface reconstruction based on image stitching from gastric endoscopic video sequence", Proceedings of SPIE Optics & Photonics 2013, Applications of Digital Image Processing XXXVI, San Diego, California, USA, vol.8856, pp.88561I-XII, 2013. [2] Ye Li, Jun Ohya, Toshio Chiba, ***Rong Xu***, and Hiromasa Yamashita, "Towards Recognizing Surgeon's Action during Suture Surgery from the Video Sequence", Proceedings of the 9th Asian Conference on Computer Aided Surgery (ACCAS 2013), Tokyo, JS1-A2, 2013. [3] Ye Li, Jun Ohya, Toshio Chiba, ***Rong Xu***, and Hiromasa Yamashita, "Recognizing surgeon's actions during suture operations from video sequences", Proceedings of SPIE Medical Imaging 2014, San Diego, California, United States, Tracking No. MI14-MI102-188, Paper No. 9034-42, accepted, 2014.
Domestic Academic Meetings (Non-Reviewed)	[1] ***Rong Xu*** and Jun Ohya, "An Improved KFCM Algorithm in the Segmentation of Brain MRIs", Proceedings of the 9th Forum on Information Technology (FIT 2010), Kyushu University, Japan, G-017, vol.2, pp.563-564, Sep. 2010. [2] ***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "A Study of Segmenting Ultrasound Image Based on Bilateral Filtering and Fuzzy Clustering", Proceedings of IEICE General Conference 2011 (EiC 2011), Tokyo City University, Japan, D-16-2, vol.2, pp.205, Mar. 2011. [3] ***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "A Study of Automated Fetal Head Detection by Pre-processing Based On Ultrasound Image Gray Feature and Iterative Randomized Hough Transform", Proceedings of the 10th Forum on Information Technology (FIT 2011), Hakodate University, Japan, G-004, vol.2, pp.529-530, Sep. 2011. [4] ***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "Improving Iterative Randomized Hough Transform for Automatic Detection of Fetal Head from Ultrasound Images", Proceedings of IEICE Technical Report 2011, Nogoya University, Japan, CS2011-76, IE2011-100, vol.111, no.349, pp.127-132, Dec. 2011. [5] ***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie,

	"A Study on 3D Fetal Model Construction with Oral Cavity and Airways", Proceedings of IEICE General Conference 2012 (EiC 2012), Okayama University, Japan, D-16-9, vol.2, pp.205, Mar. 2012. [6] ***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "A Cone Based Electromagnetic 3D Ultrasound Calibration", Proceedings of IEICE Technical Report 2012, Kobe University, Japan, vol.111, no.499, PRMU2011-281, pp.253-258, ISSN: 0913-5685, Mar. 2012. [7] ***Rong Xu***, Tomoki Hayashida, Jun Ohya, Harumi Kawamura, "Study of Estimating the Orientation of the Illumination from RGB and Depth Images", Proceedings of IIEEJ Media Computing Conference 2012, Waseda University, Japan, R7-4, pp.1-4, Jun. 2012. [8] ***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "A Registration Between 3D Ultrasound Images and 3D Fetal Model for Locating a Fetal Mouth in a Fetal Surgical Navigation System", Proceedings of the 11th Forum on Information Technology (FIT 2012), Hosei University, Japan, G-014, vol.2, pp.373-374, Sep. 2012. [9] ***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "A Study of Automatic Fetal Face and Mouth Detection From 3D Ultrasound Images for Navigating FETO Surgeries", Proceedings of IEICE Technical Report: MI 2012, Yamaguchi University, Japan, vol.112, no.271, pp.59-64, Oct. 2012. [10] ***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "A Robust ICP Algorithm for Automatic Registration of 3D Fetal Facial Surface", Proceedings of IEICE General Conference 2013 (EiC 2013), Gifu University, Japan, D-16-5, vol.2, pp.196, Mar. 2013.
Domestic Academic Meetings (Non -Reviewed, Co-author)	[1] Naotomo Tatematsu, ***Rong Xu***, Jun Ohya, Harumi Kawamura, and Shunichi Yonemura, "A Fundamental Study of Estimating Lighting Source's Position from A Single Image by Detecting Spherical Regions", Proceedings of the 10th Forum on Information Technology (FIT 2011), Hakodate University, Japan, H-049, vol.3, pp.215-216, Sep. 2012. [2] Tianliang Xie, ***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "A Study of the Detection of Fetal Mouth for An Automatic FETO Surgery", Proceedings of IEICE General Conference 2012 (EiC 2012), Okayama University, Japan, D-12-14, vol.2, pp.108, Mar. 2012. [3] Tianliang Xie, ***Rong Xu***, Jun Ohya, Bo Zhang, Yoshinobu Sato, and Masakatsu G. Fujie, "Research on Tracking Fetus' Mouth in Endoscope Video Sequence for Supporting Fetal Surgeries", Proceedings of IEICE Technical Report 2012, Kobe University, Japan, vol.111, no.499, PRMU2011-278, pp.235-240, ISSN: 0913-5685, Mar. 2012.

	[4] Mengyao Duan, ***Rong Xu***, and Jun Ohya, "A Study on 3D Point Cloud Stitching for SFM Reconstruction" , Proceedings of IEICE General Conference 2013 (EiC 2013), Gifu University, Japan, D-11-78, vol.2, pp.78, Mar. 2013. [5] Mengyao Duan, ***Rong Xu***, and Jun Ohya, "Endoscope based gastric 3D reconstruction and surface mosaic", Proceedings of the 12th Forum on Information Technology (FIT 2013), Tottori University, Japan, H-047, vol.3, pp.197-198, Sep. 2013. [6] Mengyao Duan, ***Rong Xu***, and Jun Ohya, "Study of Structure-from-motion Based 3D Reconstruction and Surface Mosaic", Proceedings of IEICE Technical Report 2013, Tottori University, Japan, vol.113, no.196, PRMU2013-44, pp.113-118, Sep. 2013. [7] Ye Li, Jun Ohya, Toshio Chiba, ***Rong Xu***, and Hiromasa Yamashita, "Fundamental Study of Recognizing the Surgeon's Action during Suture Surgery from the Video Sequence", Proceedings of the 12th Forum on Information Technology (FIT 2013), Tottori University, Japan, H-038, vol.3, pp.179-180, Sep. 2013.
Book (Chapter)	***Rong Xu***, Limin Luo and Jun Ohya (2012). Segmentation of Brain MRI, Advances in Brain Imaging, Dr. Vikas Chaudhary (Ed.), ISBN: 978-953-307-955-4, InTech, DOI: 10.5772/27596. Available from: http://www.intechopen.com/books/advances-in -brain-imaging/segmentation-of-brain-mri