Framework for ultrasonography-based augmented reality in robotic surgery : application to transoral surgery and gastrointestinal surgery

Jun Shen

Jun Shen. Framework for ultrasonography-based augmented reality in robotic surgery : application to transoral surgery and gastrointestinal surgery. Bioengineering. Université de Rennes, 2019. English. NNT : 2019REN1S078 . tel-02570502

HAL Id: tel-02570502
https://theses.hal.science/tel-02570502
Submitted on 12 May 2020

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES 1
COMUE UNIVERSITÉ BRETAGNE LOIRE

ÉCOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies
de l'Information et de la Communication
Spécialité : *Signal, image, vision*

Par

Jun SHEN

Framework for Ultrasonography-based Augmented Reality in Robotic Surgery : Application to Transoral Surgery and Gastrointestinal Surgery

Thèse présentée et soutenue à **Rennes**, le **11 Décembre 2019**

Unité de recherche :
Laboratoire Traitement du Signal et de l'Image (LTSI) et
Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM)

Composition du Jury :

Président :	Laurent SARRY	Professeur, Institut Pascal - Université Clermont Auvergne
Rapporteurs :	Marie-Aude VITRANI	Maîtresse de Conférences, ISIR, Sorbonne université
	Emmanuel PROMAYON	Professeur, TIMC, Université Grenoble-Alpes
Examinateurs :	Nabil ZEMITI	Maître à Conférences, LIRMM, Université de Montpellier
Dir. de thèse :	Jean-Louis DILLENSEGER	Maître de Conférences, LTSI, Université de Rennes 1
Co-dir. de thèse :	Philippe POIGNET	Professeur, LIRMM, Université de Montpellier

Acknowledgement

Firstly, I would like to thank my thesis committee: Prof. Laurent SARRY, Dr. Marie-Aude VITRANI and Prof. Emmanuel PROMAYON for their insightful comments and encouragement. I would also like to express my sincere gratitude to my supervisors Dr. Nabil ZEMITI, Dr. Jean-Louis DILLENSEGER and Prof. Philippe POIGNET for their continuous support of my research work through each stage of the process. Many thanks to Nabil and Jean-Louis, your guidance helped me all the time of the research and the writing of this thesis.

I would like to thank my colleagues and friends from LIRMM and LTSI who have contributed immensely to my personal and professional time. Many thanks to our research engineers – Oscar, Auguste, Guillaume – and surgeons – Christophe, Laurent, Agnès. Without your valuable knowledge and professional support, I could not finish this work. I would also like to greatly thank my colleagues and friends Oscar, Alonso, Fabien, Nicolas (Nicoco and Marlène), Quentin, Xixi, Jing, Liuyang, Jie, David, Momo Boushaki and Maxence who helped me to know and integrate in the french way of life when I arrived in France. You are such cool friends. Many thanks to Hélène, Pablo, Eugenia, Karim, Wentao, Xu, Nicolas, Elena and Axel with whom I developed warm friendship in the lab of LTSI. I will miss our joint lunches, the great trip in Paris and a bunch of laughs and jokes. I would like to say a special thank you to Hélène. I appreciate all your help since I arrived in Rennes and I always know that I can trust you. I want to also say a big thank you to Lucas, Mouné, Gustavo, Pierre, Julie, Jr, Joao, Vinayak, Maxime and others for the excellent working atmosphere and many positive vibes when I came back to LIRMM.

Many thanks to my friends Yan Yijia, Wang Yingjing (A-san), Wu Tianyu (Yeme) who are always there listening to me, caring me and supporting me. Thanks to Xu Laoshi for giving me the internship opportunity to go to Germany for the very first time. Without it I would not continue my study in different countries and have such amazing experiences in my life. I want to also thank my friends Chen Lei, Shufen, Xiaoge, Junyang, Jiaru, Iris, Burcu, Daffany, Momo, Pui, Xiuming and Zhang Bo whom I met in Munich. It was because of you, I never felt lonely in a foreign city during my Master's program.

I greatly thank my French family who received me so warmly: Didier, Carole, Perrine, Paul, Roger, Jacqueline and Gisèle. I can not thank enough my family and relatives in China who spared no effort in helping my parents and me when my father had an accident last year. I want to tell my uncles, aunts and cousins a big heartfelt thank you.

Words are certainly not enough to express my gratitude to my parents who give me infinite love and endless help and raised me with love of science and freedom. Thank you Mom and Dad for everything you have done for me!

Lastly, I want to say a heartfelt thank you to François. Thank you for always being supportive, encouraging and patient to me. Because of you, I discovered a lot of wonderful and beautiful things in my life. Merci à toi d'avoir été mon rayon de soleil aussi; merci pour tous les bons moments passés, et pour tous ceux à venir.

Table of contents

Résumé

Contents

1 Introduction

Les travaux de cette thèse s'inscrivent dans le cadre de la chirurgie assistée par ordinateur et ont été financés par le Labex CAMI [1] et la Région Bretagne.

Ce travail de recherche a été encadré Nabil ZEMITI (université de Montpellier) et supervisé par Jean-Louis DILLENSEGER (Université de Rennes 1) et Philippe POIGNET (université de Montpellier) et s'est déroulé au sein de l'équipe IMPACT du Laboratoire

---

[1] Computer Assisted Medical Interventions : http ://cami-labex.fr/

Traitement du Signal et de l'Image (LTSI, Rennes) et de l'équipe DEXTER du Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM).

La chirurgie robotisée fait partie des chirurgies à accès minimal avec ceci de spécifique que ce sont des bras robotisés qui portent la caméra endoscopique et les instruments chirurgicaux et que le chirurgien les contrôle à l'aide d'une console de commande et de visualisation. Cette chirurgie a de nombreux avantages en termes de précision du geste et d'accès à des zones non atteignables par chirurgie classique. Elle présente pourtant certains inconvénients pour la localisation des tumeurs dans le champ opératoire : la tumeur est souvent invisible à la vue endoscopique car cachée sous la surface des organes à traiter; une palpation manuelle pour localiser la tumeur est également impossible; de plus les instruments robotisés n'ont pas de retour haptique. La solution pour le chirurgien est alors de se former une image mentale de la tumeur à partir de l'image préopératoire utilisée pour le planning (scanner X ou IRM) mais les déformations induites par le geste (mise en place de rétracteurs, gonflement de l'abdomen par du gaz, ...) empêchent un bon recalage de l'image mentale vers le patient. Il est également possible d'utiliser des images peropératoires telles la fluoroscopie, le cone beam CT, l'IRM ouverte ou l'échographie pour localiser la tumeur en temps réel. La question se pose alors de comment transmettre cette information au chirurgien qui a la tête dans la console ? Une solution est alors la réalité augmentée. Dans cette thèse nous allons développer une solution de réalité augmentée qui, à partir d'une échographie 3D peropératoire, va extraire la tumeur et projeter cette information sur la vue endoscopique stéréoscopique du chirurgien. Pour cela nous allons devoir lever plusieurs verrous technologiques comme la calibration de l'échographe 3D et celle de l'endoscope, la localisation et le suivi de ces deux outils d'acquisition, et l'adaptation de ces outils à de vraies problématiques médicales. En partenariat avec des équipes médicales, nous allons adapter et évaluer cette chaîne de réalité augmentée pour la résection de tumeurs de base de langue par chirurgie transorale robotisée et pour des tumeurs du bas rectum par chirurgie laparoscopique robotisée.

2 Contexte médical

Ce chapitre introduit les deux applications médicales en chirurgie robotisée sur lesquelles nous avons élaboré notre solution de réalité augmentée. Les deux applications médicales ont des problématiques communes en termes d'impossibilité de localiser les tumeurs par la seule vision endoscopique et la possibilité de retrouver cette information à l'aide d'une échographie peropératoire.

2.1 Chirurgie transorale robotisée pour le traitement des tumeurs de base de langue

Les tumeurs de base de langue affectent directement les fonctions de la respiration, de la parole et de la déglutition des patients. Une des seules solutions est alors l'intervention chirurgicale. Toutefois la base de langue est difficile à atteindre et peut nécessiter une incision des muscles sus-hyoïdiens. La chirurgie transorale robotisée est tout indiquée pour atteindre directement la base de langue par voie purement orale sans autre geste complémentaire. Le protocole de traitement est le suivant : dans un premier temps, le chirurgien localise la tumeur et son extension par palpation, il se forge alors une image mentale de la tumeur. Il dégage ensuite la voie orale en plaçant un écarteur sur la langue. Il place alors les manipulateurs et l'endoscope du robot dans cet espace pour réséquer la tumeur (figure 1). Les challenges qui se posent aux chirurgiens sont alors les suivants : le chirurgien ne voit pas la tumeur sur la vue endoscopique, il perd tout retour haptique des instruments et il est obligé de réséquer en suivant sa carte mentale de la scène et des déformations qu'il y induit lors du geste.

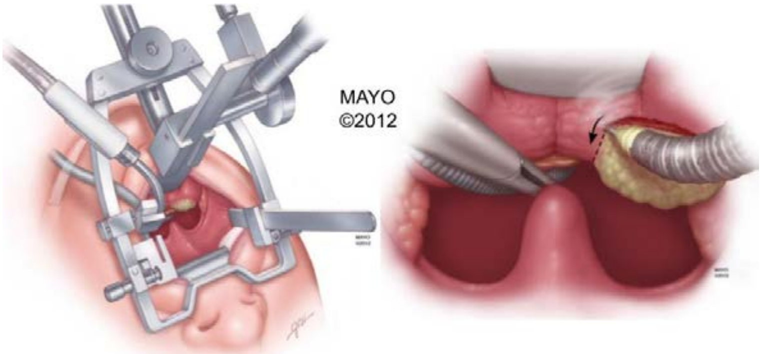


FIGURE 1 – Chirurgie transorale robotisée des tumeurs de base de langue. Source : [Van Abel and Moore, 2013].

Du fait de l'énorme déformation générée par l'écarteur, il semble extrêmement difficile de reprojeter par recalage une information de l'imagerie de diagnostic (scanner X ou IRM) sur la vue endoscopique. Une solution est dès lors d'utiliser une imagerie peropératoire afin de localiser la tumeur lors du geste. L'échographie 3D par voie trans-cervicale est une bonne voie d'abord pour imager la base de langue et permettre la localisation, la délimitation et le suivi de la tumeur en temps réel.

2.2 Chirurgie laparoscopique robotisée des cancers du bas rectum

Les cancers colorectaux sont parmi les plus fréquents et sont la deuxième cause de mortalité par cancer [Bray et al., 2018]. Concernant les tumeurs rectales, en fonction de leur stade, de leur localisation et de leur extension, une chirurgie conservatrice peut être envisagée. Dans le cas des tumeurs du bas rectum, l'exégèse peut être réalisée par voie laparoscopique robotisée. Si les extensions de la tumeur se trouvent à plus de 10 mm du canal anal, le geste est alors le suivant. Par voie laparoscopique, le chirurgien pratique la résection de la zone tumorale en prenant une marge de 10 mm dans la zone distale par rapports à l'extension de la tumeur (figure 2). Cette marge permet ainsi de préserver les fonctions du canal anal. La grande difficulté rencontrée lors de la chirurgie robotisée est que la tumeur n'est pas visible sur les vues endoscopiques. Il est alors extrêmement difficile de définir cette marge distale à 10 mm, et donc de préserver au mieux les sphincters du patient. Actuellement, lors de l'intervention robotisée, un second chirurgien estime la position de la tumeur par toucher rectal. L'information retournée par ce chirurgien est alors utilisée pour marquer la marge distale de résection sur la paroi du rectum (à l'aide du bistouri électrique par voie laparoscopique). Toutefois l'estimation de cette marge est peu précise, car elle dépend de la sensibilité du second chirurgien à estimer l'extension distale de la tumeur et à transmettre cette information au chirurgien maniant le robot.

Une solution est dès lors d'utiliser une imagerie peropératoire afin de localiser l'extension de la tumeur et de reprojeter cette information, avec si possible la marge distale de résection, sur la vue laparoscopique du chirurgien. Ce dernier pourra alors marquer cette marge sur la paroi du rectum avant de procéder à l'ablation. L'échographie transrectale 3D semble être la modalité d'acquisition la plus adaptée à cette chirurgie.

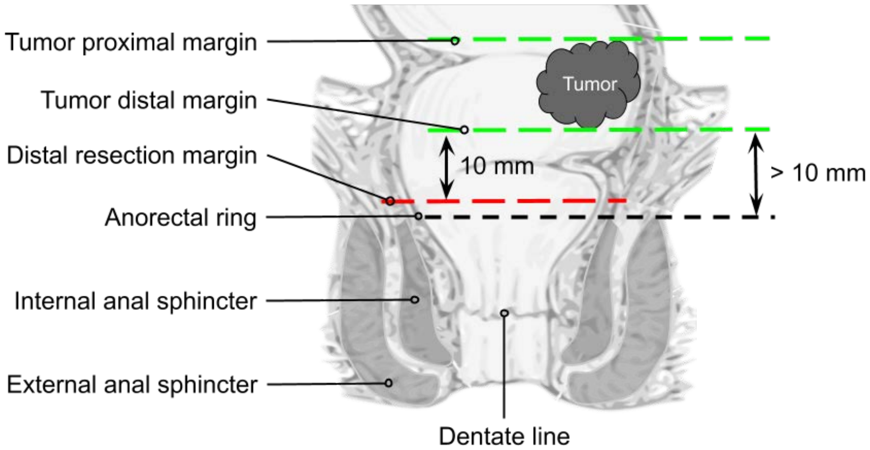


FIGURE 2 – Localisation de la tumeur à plus de 10 mm du canal anal et marges de résection envisagées. Figure reproduite de Anorectum-en.svg (auteur : National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)), sous licence Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).

3 Échographie peropératoire et réalité augmentée pour le guidage d'un geste

Ce chapitre présente un état de l'art sur les techniques de réalité augmentée en chirurgie minimalement invasive et plus particulièrement sur le report d'une information acquise par échographie peropératoire dans le champ de vue endoscopique du médecin. Cet état de l'art va permettre de justifier les choix faits durant le travail de Thèse.

De manière plus globale, la réalité augmentée consiste à superposer des objets virtuels portant une information dans l'environnement réel du chirurgie, et ceci en temps réel durant la réalisation du geste. Dans le cadre plus spécifique de la chirurgie mini-invasive et plus particulièrement de la chirurgie robotisée, l'environnement réel est visualisé à l'aide de caméras endoscopiques. La procédure de réalité augmentée consiste alors à (figure 3) :

1. générer un objet virtuel à partir de l'imagerie pré ou peropératoire, de la localisation de cette imagerie et de l'information de planning (par exemple les marges de résection) ;
2. de capturer en temps réel la vue du champ opératoire à l'aide des caméras endoscopiques ;

3. de recaler en temps réel l'objet virtuel dans l'espace de la vue endoscopique ;
4. de superposer sur la vue endoscopique une vue de l'objet virtuel afin de fournir cette information complémentaire au chirurgien.

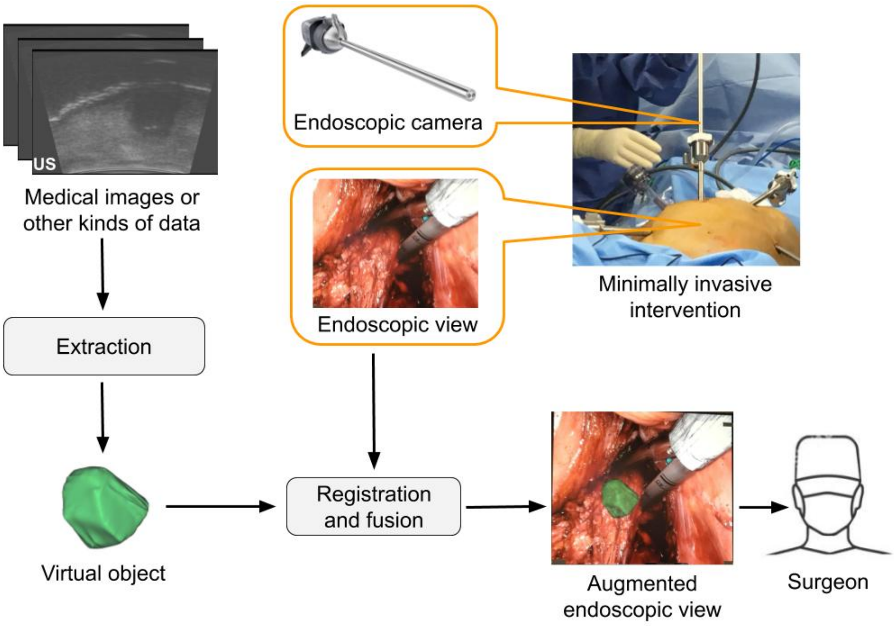


FIGURE 3 – Élément de l'implémentation d'une chaîne de réalité augmentée en chirurgie à accès minimal.

Dans le cadre de nos applications cliniques, les objets virtuels sont la tumeur et les marges de résection. La tumeur est visible dans l'échographie peropératoire. La segmentation de cette tumeur est hors de propos de notre Thèse. Par contre, nous tâcherons de répondre aux questions suivantes (nos contributions seront mises en italique).

- Comment localiser la pose de l'échographe et de l'endoscope 3D dans le champ opératoire ? Au vu de la littérature et des solutions techniques existantes, le tracking optique nous paraît la meilleure solution. Pour cela des marqueurs optiques seront placés sur l'échographe et l'endoscope afin d'estimer leur pose à l'aide d'un localisateur.
- Comment calibrer l'échographe 3D ? C'est-à-dire, comment estimer la relation spatiale (la transformation géométrique) entre le volume acquis par l'échographe et le marqueur optique placé sur lui ? Les solutions proposées dans la littérature sont généralement compliquées à mettre en œuvre.

Nous pensons qu'une solution plus simple est nécessaire pour être utilisable dans la salle d'opération. Cette solution fait l'objet d'une des contributions de notre Thèse.

- Comment calibrer la vue de l'endoscope 3D par rapport au marqueur optique placé sur lui ? Cette problématique a déjà été bien étudiée dans la littérature. La méthode dite du "hand-eye" est parfaitement adaptée à notre problématique. Nous l'utiliserons en l'adaptant toutefois aux objectifs grand-angle de notre endoscope stéréo.
- Comment associer ces différents éléments ? Et quelle est l'influence de chacun de ces éléments sur la précision de la chaîne totale de réalité augmentée ?

  Afin de répondre à cette question, nous proposerons une évaluation de la précision de la calibration de l'échographe, de celle de l'endoscope et de la chaîne complète.
- Notre chaîne de réalité augmentée est-elle bien adaptée à nos problématiques médicales ?

  Deux expérimentations sur fantômes physiques ont été mises en place afin de démontrer l'utilisation de notre chaîne de réalité augmentée dans les conditions les plus proches possibles des vraies interventions : la chirurgie robotisée des tumeurs de base de langue et la chirurgie robotisée des tumeurs du bas rectum.

4 Implémentation des éléments de la chaîne de réalité augmentée

4.1 Présentation des éléments de la chaîne de réalité augmentée

Afin de définir les éléments de notre chaîne de réalité augmentée, nous allons dans un premier temps analyser la chaîne de traitement de l'information à fournir au chirurgien (en violet dans la figure 4) :

- l'échographie 3D est l'imagerie peropératoire qui va fournir l'information cachée nécessaire au chirurgien ;
- le modèle virtuel de la tumeur est extrait par segmentation du volume échographique. Une information supplémentaire, comme les marges de résection, peut enrichir cet objet virtuel. Comme il a été notifié précédemment, cette étape de segmentation en

temps réel de la tumeur n'a pas fait l'objet de travaux durant notre Thèse. Le choix de la méthode la plus adaptée aux données est laissé aux futurs utilisateurs ;

- en parallèle à l'imagerie peropératoire, le chirurgien visualise la scène opératoire au travers d'une caméra endoscopique stéréoscopique portée par le robot ;
- le modèle virtuel est alors recalé dans l'espace de l'endoscope 3D et est superposé à la vue capturée par cet endoscope ;
- cette vue augmentée est alors fournie au chirurgien par l'intermédiaire du système de vision binoculaire de la station de contrôle du robot (dans notre cas un casque "Head-Mounted Display" - HMD).

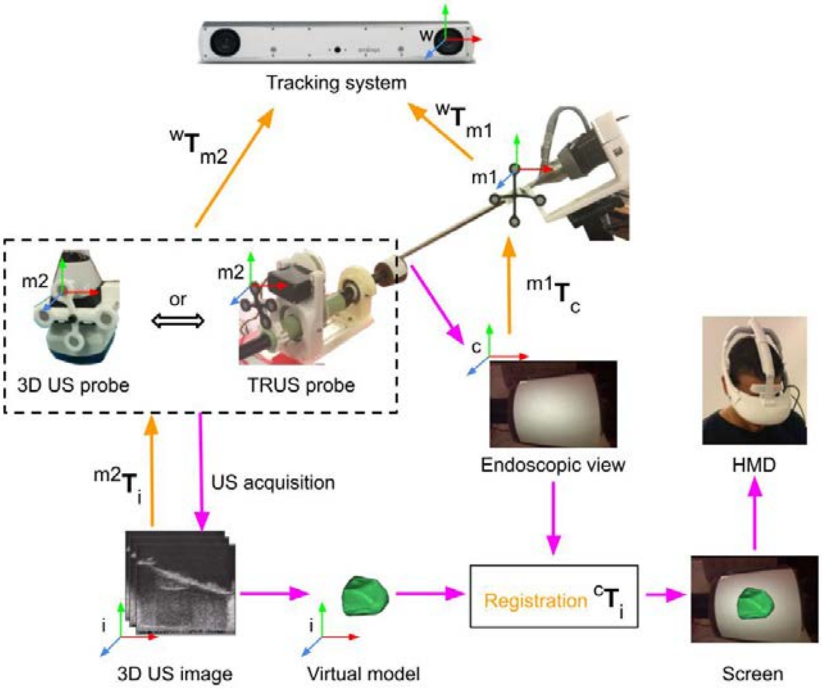


FIGURE 4 – Éléments de la chaîne de réalité augmentée. En violet, la chaîne de traitement de l'information. En orange, la chaîne de recalages et de calibrations de capteurs.

Cette chaîne de traitement de l'information nécessite en parallèle toute une chaîne de recalages d'images et de calibrations de capteurs. Cette chaîne est représentée en orange dans la figure 4. Dans cette figure et également dans la suite du manuscrit, nous avons défini plusieurs référentiels spatiaux (systèmes de coordonnées) liés aux différents objets : i celui de l'image échographique 3D, $m1$ lié à un marqueur optique fixé solidement sur l'endoscope 3D, $m2$ lié à un marqueur optique fixé solidement sur l'échographe 3D, c lié au volume vu par l'endoscope 3D et w le repère monde. Nous notons également ${}^{b}\mathbf{T}_{a}$ la transformation homogène du repère *a* vers le repère *b*. La chaîne de recalages et de calibrations de capteurs comporte :

- La localisation de la pose de l'échographe et de l'endoscope 3D dans le champ opératoire. Pour cela nous utilisons un système de localisation optique qui définit le repère monde w avec deux marqueurs fixés solidement sur l'endoscope (marqueur $m1$) et sur l'échographe (capteur $m2$). Le localisateur permet d'estimer ${}^{w}\mathbf{T}_{m1}$ et ${}^{w}\mathbf{T}_{m2}$.
- La calibration de l'échographe 3D permet d'obtenir la relation ${}^{m2}\mathbf{T}_{i}$. La méthode de calibration que nous avons élaborée et son évaluation seront présentées dans la section **Calibration de l'échographe 3D**.
- La calibration de la caméra stéréoscopique de l'endoscope nous donne ${}^{m1}\mathbf{T}_{c}$. Nous avons utilisé la méthode "hand-eye" (section **Calibration de l'Endoscope 3D**).

Nous pouvons alors définir la transformation géométrique homogène de l'image échographique 3D vers le l'espace 3D de l'endoscope par :

$${}^{c}\mathbf{T}_{i} = ({}^{m1}\mathbf{T}_{c})^{-1} \cdot ({}^{w}\mathbf{T}_{m1})^{-1} \cdot {}^{w}\mathbf{T}_{m2} \cdot {}^{m2}\mathbf{T}_{i}$$

4.2 Calibration de l'échographe 3D

L'objectif de la calibration est de trouver la relation entre le volume acquis par l'échographique et le monde réel. Plus particulièrement dans notre cas, le monde réel est le marqueur optique $m2$ fixé sur l'échographe. La plupart des méthodes de calibration utilisent des fantômes de calibration visibles en échographie (des cadres 3D avec des fils [Prager bermeier] ; des points sur un plan [detmer Hsu] ; ou des plans perpendiculaires [Rousseau, Abeuysekera]). L'idée générale de ce type de calibration est la suivante : 1) dans un premier temps, un marqueur de localisation est placé sur le fantôme et le fantôme est alors calibré par rapport à ce marqueur (pour trouver la relation entre les points remarquables du fantôme et le repère monde) ; 2) un second marqueur de localisation est placé sur l'échographe 3D ; 3) le fantôme est imagé, les amers remarquables du fantôme sont segmentés et la position du référentiel du fantôme est estimée dans le repère image à partir des amers segmentés ; 4) la calibration se fait alors en associant cette position et les poses du fantôme et de l'échographe estimées à l'aide du localisateur externe.

Plusieurs points de cette méthode sont sources de difficultés de manipulation ou sont sujets à des imprécisions. Les fantômes proposés dans la littérature sont difficilement exploitables dans notre cas car ils sont artisanaux, assez compliqués à fabriquer et à mettre en œuvre et peu reproductibles. La calibration du fantôme par rapport au marqueur de localisation est souvent difficile et source d'imprécision. La localisation de ce marqueur est souvent délicate car le fantôme est placé dans de l'eau. En fonction du type d'amers remarquables du fantôme, la segmentation de l'image échographique peut être assez imprécise et donc il peut être difficile d'estimer précisément dans l'image la position du référentiel du fantôme à partir des amers segmentés.

Afin de contrebalancer ces imprécisions, nous proposons d'utiliser un nouveau type de fantôme imaginé en Conception Assistée par Ordinateur (CAO, Computer Aided Design, CAD en anglais) et facile à fabriquer à l'aide d'une imprimante 3D, de géométrie relativement simple, facilement segmentable dans l'image échographique et évitant si possible l'étape de calibration du fantôme. Ce nouveau fantôme doit permettre de simplifier et d'automatiser au maximum l'étape de calibration afin de la rendre accessible à un non-spécialiste dans le bloc opératoire.

Solution pour la sonde 3D utilisée pour la chirurgie de base de langue.

Dans un premier temps, partant du modèle CAO de la coque de l'échographe, nous avons conçu en CAO un support permettant de fixer solidement et solidariser le marqueur de localisation $m2$ à la coque de l'échographe.

Le fantôme de calibration a lui aussi été conçu en CAO. Il est globalement composé de deux pièces (figure 5) : A) une partie qui sera imagée par l'échographe, qui se présente sous la forme d'une boîte ouverte, de taille $25 \times 20 \times 10$ mm, ajourée par quatre trous de rayons 4, 3, 2 et 2 mm ; B) une partie que sera fixée de manière précise, soit directement sur le marqueur de localisation $m2$, soit sur son support, solidarisant ainsi le fantôme à l'échographe. Cette partie permet de faire coïncider mécaniquement le référentiel $m2$ de l'échographe et le référentiel du fantôme p. La transformation ${}^{m2}\mathbf{T}_p$ entre le fantôme p et le marqueur $m2$ est donc directement donnée par le modèle CAO du fantôme.

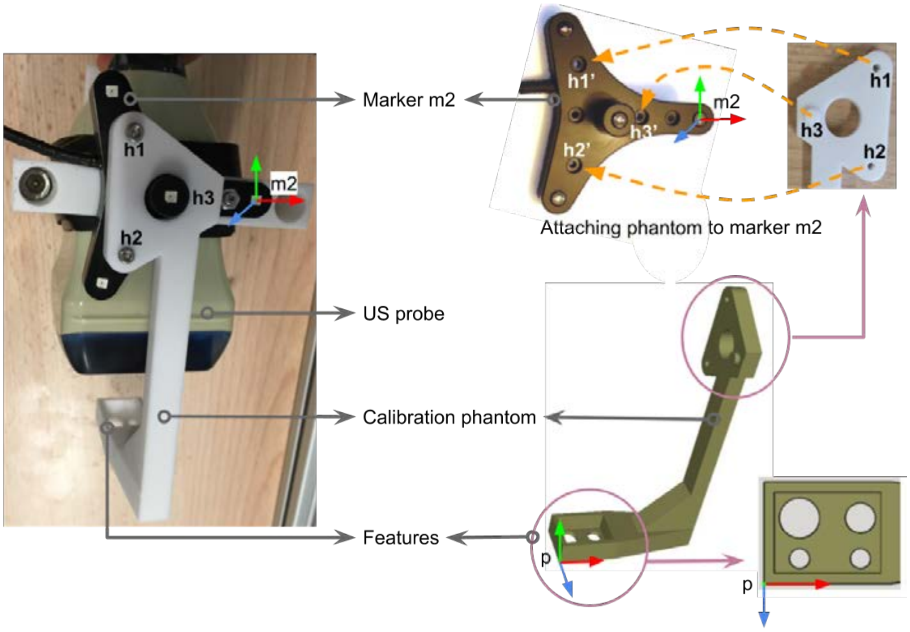


FIGURE 5 – Fantôme de calibration.

Le processus de calibration est alors le suivant (figure 6) : 1) le fantôme de calibration est fixé sur l'échographe ; 2) une image échographique 3D de ce fantôme est acquise dans de l'eau ; 3) l'image échographique est segmentée automatiquement (extraction des gradients et transformée de Hough sur l'image de gradients) ; 4) un maillage est généré sur l'objet segmenté ; 5) ce maillage est recalé sur le modèle CAD du fantôme donnant ainsi facilement la transformation ${}^{p}\hat{\mathbf{T}}_i$; et 6) l'équation finale de calibration ${}^{m2}\hat{\mathbf{T}}_i$ est alors donnée par ${}^{m2}\mathbf{T}_p \cdot {}^{p}\hat{\mathbf{T}}_i$.

Le fantôme de calibration peut alors être enlevé de l'échographe sans toucher au marqueur $m2$.

Afin d'évaluer la précision de la calibration, nous avons calculé la "Fiducial Registration Error" (FRE) pour certains points remarquables du fantôme. Ces points sont dans un premier temps pointés manuellement sur l'image échographique du fantôme puis reprojetés vers le modèle CAO à l'aide de ${}^{m2}\hat{\mathbf{T}}_i$. La FRE est la distance entre le point issu de l'échographie reprojeté et sa vraie position dans le modèle CAO. La moyenne des FRE est de l'ordre de 0,36 mm.

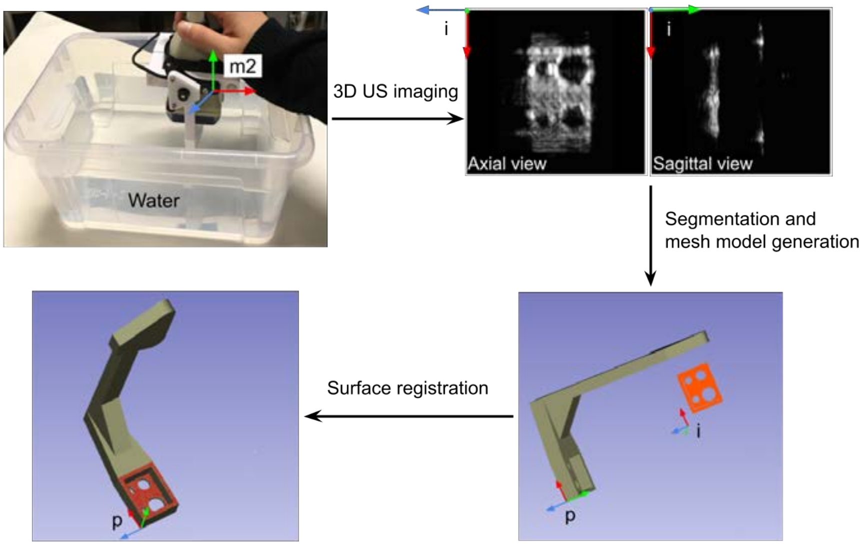


FIGURE 6 – Processus de calibration de l'échographe.

Une fois le fantôme enlevé, nous avons également imagé, à l'aide d'un échographe, le bout d'un stylet portant un marqueur de localisation. L'erreur quadratique moyenne entre la position réelle du bout du stylet et celle obtenue par l'image échographique calibrée est de 1,39 mm (avec toutefois une imprécision annoncée de l'ordre de 0,9 mm pour la localisation réelle du bout du stylet par le système de localisation). L'erreur quadratique moyenne que nous obtenons est inférieure à celles annoncées par les méthodes de la littérature.

Solution pour la sonde transrectale utilisée pour la chirurgie des tumeurs du bas rectum

L'imagerie peropératoire prévue est l'échographie transrectale (TRUS) 3D. Nous n'avions pas encore de sondes de ce type à notre disposition, nous avons donc adapté une sonde TRUS 2D en la motorisant et en la localisant spatialement à l'aide du marqueur $m2$ afin de réaliser une sonde 3D (figure 7).

Le fantôme de calibration a été adapté à la morphologie de la sonde (figure 7). Nous lui avons donné une forme cylindrique avec comme amers remarquables 3 trous de rayons 4, 3 et 2 mm et un rectangle de 4×15 mm. Comme la sonde échographique est mobile du fait de la motorisation, nous n'avons pas pu lier mécaniquement le fantôme au marqueur $m2$.

Nous avons donc attaché un marqueur de localisation $m3$ au fantôme. La relation entre $m3$ et $m2$ est alors estimée par un localisateur externe (à terme, pour des sondes TRUS 3D, nous pensons pouvoir lier mécaniquement le fantôme à la sonde, comme pour la sonde utilisée pour la base de langue).

La calibration suit la méthode décrite pour la sonde de base de langue avec comme point central le recalage du maillage du fantôme segmenté de l'image échographique avec le modèle CAO du fantôme.

Après calibration et démontage du fantôme, nous avons mesuré la précision de la calibration à l'aide du stylet portant un marqueur de localisation. L'erreur quadratique moyenne mesurée sur cinq positions dans l'espace entre la position réelle du bout du stylet et celle obtenue par l'image échographique calibrée est de 0,9 mm.

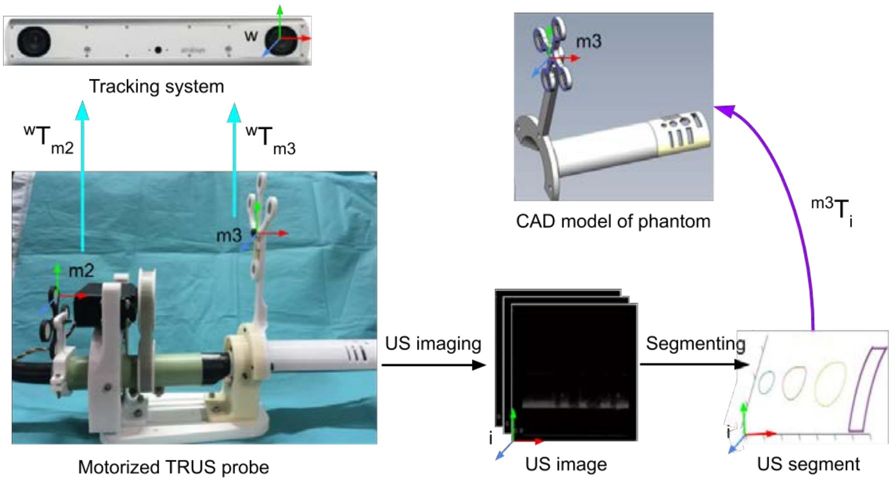


FIGURE 7 – Processus de calibration de l'échographe transrectal.

4.3 Calibration de l'endoscope 3D

L'objectif de la calibration est le même que pour l'échographe 3D, c'est-à-dire de trouver la relation entre le monde réel (représenté par le marqueur de localisation $m1$ fixé sur l'endoscope) et l'espace 3D vu par les caméras stéréoscopiques. Pour cela, nous avons simplement adapté des solutions classiques à notre problématique.

D'une part, nous avons calibré le système stéréoscopique (paramètres intrinsèques, extrinsèques et distorsions) à l'aide d'un damier adapté au champ de vue de l'endoscope (faible distance et objectifs grand angle) et en utilisant les outils de la bibliothèque OpenCV.

D'autres part, la calibration entre la vue stéréoscopique et le marqueur $m1$ a été réalisée à l'aide de la méthode dite "hand-eye", largement utilisée en robotique (figure 8) [Tsai]. L'idée est d'imager un objet, dont les caractéristiques spatiales sont connues (dans notre cas le damier utilisé pour la calibration stéréo sur lequel est positionné un marqueur $m4$), selon plusieurs incidences. Connaissant les poses de l'objet de calibration et de l'endoscope aux différentes incidences, il est possible d'estimer la relation ${}^{m1}\mathbf{T}_c$ entre le point de vue stéréoscopique c et le marqueur $m1$.

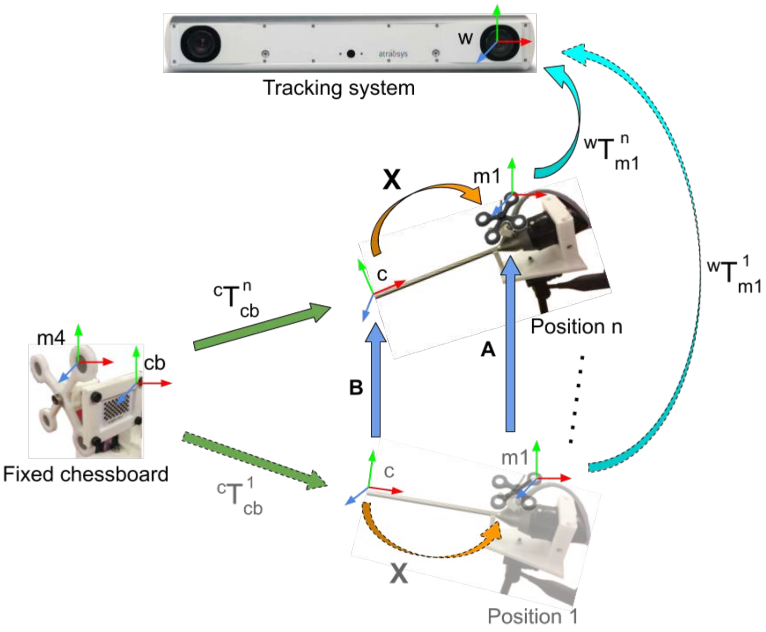


FIGURE 8 – Principe de la calibration "hand-eye".

Cette calibration a été évaluée à l'aide d'un disque réflectif dont la pose est estimée à l'aide d'un marqueur de localisation. La précision de la calibration a été estimée en projetant le contour 3D connu du disque sur les images stéréoscopiques et en mesurant la distance entre le disque projeté et les contours de ce disque sur les images stéréoscopiques. Nous avons constaté une erreur quadratique moyenne de l'ordre de 0,5 mm lorsque le disque est dans le champ de vue moyen de l'endoscope et de 1,5 mm lorsque l'on s'éloigne de ce champ de vue moyen (trop proche, trop loin, ou en limite angulaire).

4.4 Évaluation de la chaîne complète de réalité augmentée

Les différents éléments de la chaîne de réalité augmentée ont été associés entre eux avec l'échographe calibré (${}^{m2}\mathbf{T}_i$), l'endoscope calibré (${}^{m1}\mathbf{T}_c$) et le système de localisation (${}^{w}\mathbf{T}_{m1}$ et ${}^{w}\mathbf{T}_{m2}$).

Pour l'évaluation, nous avons : 1) placé le fantôme de calibration sur la sonde échographique ; 2) acquis l'image échographique du fantôme ; 3) segmenté les contours des amers particuliers du fantôme (cercles et rectangles) ; 4) projeté ces contours sur les vues stéréoscopiques de l'endoscope braqué sur le fantôme à l'aide des transformations géométriques issues de la calibration et du tracking ; 5) mesuré les distances entre les contours échographiques projetés et les contours réels vus dans les images endoscopiques (figure 9).

L'erreur quadratique moyenne est de l'ordre de 0,51 mm pour la vue gauche et 0,87 mm pour la vue droite pour la sonde échographique utilisée pour la base de langue et de 0,45 mm pour la vue gauche et 0,8 mm pour la vue droite pour la sonde échographique transrectale.

Selon nos partenaires médicaux, la précision atteinte semble tout à fait acceptable.

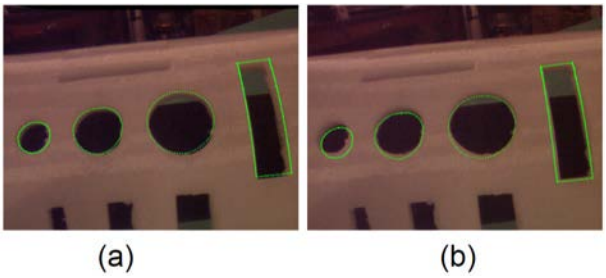


FIGURE 9 – Évaluation de la chaîne de réalité augmentée. Projection de l'information acquise par l'échographe sur les vues droite et gauche de l'endoscope.

5 Expérimentations de la chaîne de réalité augmentée sur fantômes physiques

Dans ce chapitre, nous allons présenter les 3 expérimentations de mise en œuvre et de validation de la chaîne de réalité augmentée sur fantômes physiques. La première expérimentation était assez générique et avait pour ambition de montrer la preuve de concept de l'utilisation de la chaîne de réalité augmentée pour la localisation d'une cible cachée dans un fantôme physique. Les deux autres expérimentations devaient simuler de manière la plus réaliste possible l'utilisation de la chaîne de réalité augmentée dans les deux applications médicales cibles, la chirurgie transorale robotisée de la base de langue et la chirurgie laparoscopique robotisée des tumeurs du bas rectum. Dans le cadre de la chirurgie transorale robotisée de la base de langue, le dispositif a été testé sur une langue de mouton *ex-vivo*. L'évaluation du dispositif pour la chirurgie laparoscopique robotisée des tumeurs du bas rectum a été menée sur un fantôme physique réaliste en gel. Ces deux expérimentations ont été montées en collaboration avec nos partenaires médicaux et leur ont permis d'imaginer l'apport du dispositif dans leur pratique future.

5.1 Preuve de concept de la chaîne de réalité augmentée pour la localisation d'une cible cachée dans un fantôme physique

Comme mentionné précédemment, l'ambition de cette expérimentation était surtout de démontrer la faisabilité de l'utilisation d'une chaîne de réalité augmentée pour la localisation d'une cible cachée.

Sans pertes de généricité, le *modus operandi* classique du chirurgien pour la résection d'une tumeur cachée est le suivant, 1) localisation de la tumeur sur une imagerie diagnostique préopératoire (scanner X ou IRM) ; 2) palpation de la zone tumorale avant le geste afin de localiser physiquement cette tumeur dans l'organe à traiter et 3) résection de cette zone en s'aidant de l'image mentale de la tumeur issue de l'imagerie préopératoire et de la palpation.

L'objectif (et donc le scénario) de cette première expérimentation était de montrer qu'avec une chaîne de réalité augmentée, il était possible d'avoir des résultats de localisation similaires (voire meilleurs si possible) à ceux effectués par le chirurgien de manière classique.

Pour cela nous avons fabriqué un fantôme souple (pour permettre la palpation) en utilisant du silicone (figure 10). Le fantôme est constitué d'un objet triangulaire assez dense (le modèle de tumeur) inclus dans un substrat opaque moins dense simulant le tissu humain. Le modèle de tumeur est papable de la surface du substrat. Afin de faire varier l'information palpable, nous avons créé trois fantômes en faisant varier la position de la tumeur par rapport à la surface du substrat : *A* proche de la surface, *B* au milieu et *C* en profondeur.

(a) Silicone object mimicking a soft-tissue tumor (b) Silicone phantom (c) Phantom placed in a box

FIGURE 10 – Fantôme de tumeur cachée. a) tumeur souple simulée, b) inclusion dans un substrat souple pour simuler le tissus humain, c) fantôme emboîté de force (déformant le modèle de tumeur).

Ces fantômes ont ensuite été placés dans une IRM afin de créer l'image préopératoire servant au chirurgien à établir l'image mentale de la tumeur.

Chaque fantôme est alors emboîté de force dans une boite rigide déformant par la même occasion la tumeur (ceci permet de simuler les déformations subies lors du geste). Le fantôme est alors placé sous une caméra stéréoscopique [2] (formée de 2 caméras 2D) simulant le champ opératoire. L'utilisateur visualise le champ opératoire au travers d'un casque HMD simulant la vision stéréoscopique de la console de commande du robot.

La tâche demandée à l'utilisateur est alors de délinéer sur la surface du substrat la forme

---

[2] Au moment de notre première expérimentation, nous ne disposions pas encore de l'endoscope 3D. Afin de mettre en place notre chaîne de traitement, nous avons utilisé deux caméras côte à côte afin de faire une acquisition stéréoscopique de l'image. Ces caméras ont été calibrées par la méthode expliquée dans la section **Calibration de l'Endoscope 3D**. La différence principale entre cette caméra stéréoscopique et la caméra endoscopique réside dans l'objectif grand angle de l'endoscope.

de la tumeur. Pour cela, nous avons donné à l'utilisateur un stylet sur lequel nous avons placé un capteur de localisation. Ce capteur permet de capturer la trajectoire dessinée par l'utilisateur.

L'utilisateur doit estimer la forme de la tumeur selon deux protocoles :

Selon le *modus operandi* classique : a) analyse de l'IRM préopératoire par l'utilisateur, b) palpation de la surface du substrat et c) tracé des limites de la tumeur à partir de l'image mentale formée par les informations précédentes.

Ce protocole est appelé *palp* par la suite.

Par réalité augmentée à l'aide de notre solution. Plus précisément (figure 11) :

1. dans un stade préopératoire, l'objet triangulaire est segmenté par seuillage dans le volume IRM et est maillé automatiquement afin de former l'objet virtuel ; b) dans un stade peropératoire, l'échographe fait une acquisition du fantôme physique déformé dans sa boite ;
2. dans un stade peropératoire, l'échographe fait une acquisition du fantôme physique déformé dans sa boîte ;
3. un recalage élastique entre l'IRM et l'échographe permet d'estimer les déformations subies par l'objet triangulaire [3]. Cette acquisition permet d'actualiser la forme de l'objet virtuel en fonction de déformations subies ;
4. en se basant sur les calibrations des caméras et de l'échographe 3D et sur la localisation de leurs poses par le système de localisation, l'objet virtuel est reprojeté dans l'espace des caméras et son information de surface (ou de contour) est superposée sur la vue stéréoscopique du casque HMD ;
5. l'utilisateur trace les limites de la tumeur sur la surface du substrat en se servant de la vue augmentée.

Ce protocole est appelé *AR* par la suite.

Nous avons demandé à 6 participants de procéder à la délinéation des limites de l'objet triangulaire sur la surface du substrat à l'aide des deux protocoles (*palp* et *AR*), et ceci

---

[3] Le recalage a été réalisé à l'aide de la bibliothèque logicièle Elastix (modèle de transformation : B-Spline ; mesure de similarité par Information Mutuelle).

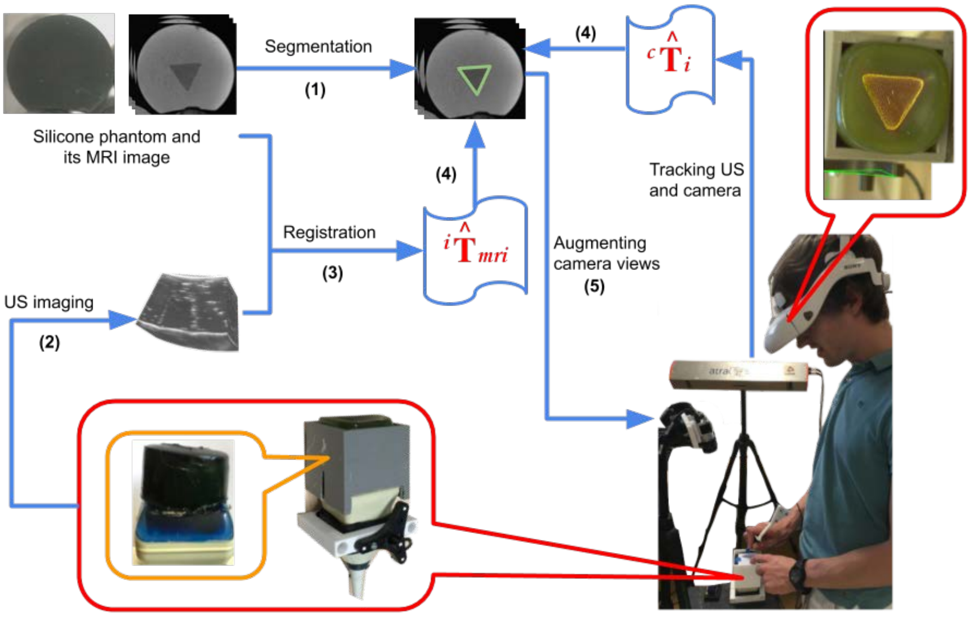


FIGURE 11 – Expérimentation pour la preuve de concept de la chaîne de réalité augmentée pour la localisation d'une cible cachée dans un fantôme physique.

pour les 3 fantômes A, B et C. La précision de la délinéation sera estimée par le coefficient de Dice (mesurant la superposition) et la distance de Hausdorff (mesure qui quantifie les dissemblances) entre la délinéation faite par l'utilisateur et la vérité terrain (la forme triangulaire projetée sur la surface).

Dans tous les cas, les utilisateurs obtenaient des meilleurs coefficients de Dice et distances de Hausdorff lorsqu'ils utilisaient la chaîne de réalité augmentée.

Nous avons également constaté une différence entre les utilisateurs. Les meilleurs scores ont été obtenus par des chirurgiens. C'était d'autant plus flagrant dans l'expérimentation sur la palpation.

Ces premiers résultats ont permis d'une part de faire une première validation de l'ensemble de la chaîne de traitement et surtout permis de démontrer les apports de la réalité augmentée pour la délinéation d'objets cachés par rapport à la technique classique utilisée par les chirurgiens (images préopératoires et palpation).

Cette expérimentation nous a également permis de recadrer notre travail. En discutant avec les chirurgiens nous avons pris conscience que l'imagerie préopératoire n'était pas forcément nécessaire pour notre chaîne de réalité augmentée. Comme les tumeurs sont visibles en échographie peropératoire, l'objet virtuel peut être extrait et modélisé à partir de cette seule modalité d'acquisition sans avoir à faire un recalage échographie/images préopératoire comme nous le pensions au début de la Thèse.

5.2 Validation de la chaîne de réalité augmentée pour la chirurgie transorale robotisée de la base de langue : expérimentation sur une langue *ex-vivo*

L'objectif de cette expérimentation était non seulement de valider nos solutions pour le guidage de résections de tumeurs de base de langue mais également de prouver au chirurgien l'utilité de l'échographie peropératoire lors de chirurgie.

L'expérimentation a été menée sur une langue de mouton *ex-vivo* d'épaisseur de 35 mm dans la section la plus large. Le modèle de tumeur de langue est le suivant : nous avons simulé une tumeur en silicone de taille approximativement $10 \times 8 \times 5$ mm et l'avons insérée dans la base de la langue au travers d'une incision, (voir figure 12-b). L'incision est ensuite suturée pour refermer la surface de la langue.

L'échographe 3D est ensuite calibré en suivant la procédure décrite dans le chapitre **Calibration de l'échographe 3D**. Du fait de la facilité de la mise en œuvre du fantôme de calibration et de l'automatisation de la procédure, le processus de calibration de la sonde prend moins de 5 min.

Dans la future intervention, l'échographie sera placée en région sous-mentonnière. Notre modèle de langue étant excisé, nous avons placé l'échographe 3D (Sonixtouch Q+ avec une sonde 4DL 14-5/38) dans la partie inférieure (opposée à la tumeur) de la langue.

La procédure de réalité augmentée est alors la suivante : la zone de la tumeur est acquise en échographie 3D et la tumeur est segmentée (manuellement dans cette expérimentation) dans le volume échographique (figure 12-a) afin de créer l'objet virtuel. La silhouette de

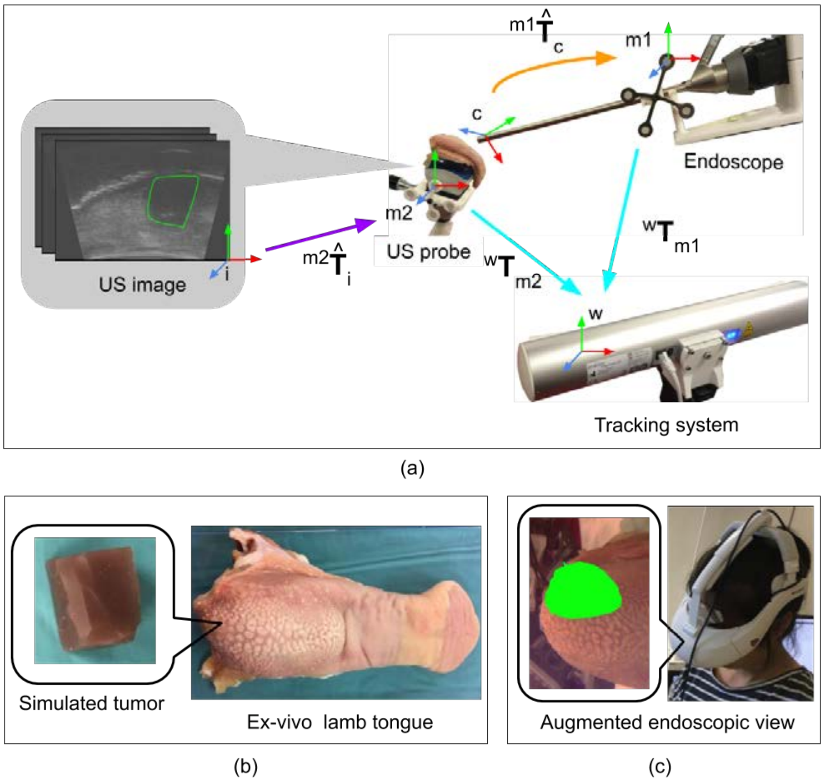


FIGURE 12 – Expérimentation sur une langue *ex-vivo*. a) chaîne de calibrations et recalages, b) modèle de tumeur de base de langue et c) vision en réalité augmentée projeté dans le casque HMD).

la tumeur est alors reprojetée sur les vues endoscopiques (figure 12-c). Le participant portant un HMD délimite, à l'aide d'un marqueur, la silhouette de la tumeur sur la surface de la langue en y incluant une marge de résections de l'ordre de 10 mm. Cette procédure se voulait être assez proche de la vraie procédure chirurgicale dans laquelle les chirurgiens marquent la zone à réséquer en cautérisant légèrement la surface de langue à l'aide du bistouri électrique placé au bout du bras manipulateur du robot chirurgical. Dans un deuxième temps, le participant a réséqué la tumeur en suivant les contours délimités en réalité augmentée.

Afin d'évaluer le succès de la résection, nous avons placé le spécimen dans de l'eau et nous l'avons imagé à l'aide de l'échographe 3D. Le volume échographique est échantillonné en 10 coupes espacées de 1 mm. Pour chaque coupe, nous avons mesuré, à l'aide du logiciel Slicer 3D, la distance minimale et maximale entre les bords de la tumeur et les bords du spécimen réséqué. La moyenne des distances maximales est de 11,7 mm et la moyenne des distances minimales est de 8,5 mm. La marge de résection semble avoir été bien respectée avec une tolérance de ± 2 mm. Cette tolérance est supérieure aux mesures d'erreurs effectuées lors du chapitre **Évaluation de la chaîne complète de réalité augmentée**. Nous supposons donc que cette imprécision vient du geste de résection lui-même et non des éléments de la chaîne.

Cette expérimentation prouve la faisabilité de la procédure sur un vrai organe. Elle nous a permis d'utiliser notre chaîne de réalité augmentée dans des conditions proches de la thérapie. Elle a montré certains points à améliorer tels la nécessité de proposer une procédure automatique de segmentation de la tumeur en image échographique ou le choix de l'information à reprojeter dans la vue binoculaire (maillage 3D ? contour de la tumeur et marges ? ...).

Les expérimentations futures vont inclure l'utilisation de bras robotisés (robot Raven) et l'évaluation du traitement d'inclusion de modèles de tumeurs plus petites (< 4mm) correspondant à des tumeurs de stade T1 ou T2.

5.3 Validation de la chaîne de réalité augmentée pour chirurgie laparoscopique robotisée des tumeurs du bas rectum : expérimentation sur un fantôme physique

Le point crucial pour la résection des tumeurs est la définition de la marge de résection distale. Cette marge doit être assez grande afin d'inclure les cellules cancéreuses diffuses mais suffisamment précise pour préserver les sphincters du patient. Nos différents échanges avec les chirurgiens ont permis d'imaginer le protocole en chirurgie robotisée laparoscopique suivant : après avoir placé le robot en condition de chirurgie laparoscopique, a) une sonde TRUS 3D est placé dans le rectum afin d'acquérir un volume de la zone candidate à la chirurgie ; b) le chirurgien segmente manuellement les contours de la tumeur (cette segmentation peut être limitée à déterminer les marges proximale et distale de la tumeur) ; c) les contours de la tumeur sont alors reprojetés dans le référentiel de la vue endoscopique à l'aide de notre chaîne de traitement. d) le système définit alors automatiquement la marge distale de résection et superpose les contours de la tumeur et le trait correspondant à la marge de résection sur la vue laparoscopique ; e) le chirurgien marque alors en laparoscopie la marge de résection sur le bas-rectum par cautérisation légère avec le bistouri électrique ; f) la sonde TRUS est alors retirée du patient et le chirurgien coupe le rectum au niveau de la marge de résection marquée par le bistouri puis procède de manière classique à la résection de la zone tumorale.

L'expérimentation que nous avons menée reproduit au mieux ce protocole chirurgical. Dans un premier temps, nous avons fabriqué un fantôme physique simulant le rectum et le colon sous la forme d'un cylindre opaque en silicone dans lequel nous avons inclus un modèle de tumeur en silicone de densité supérieure et de taille de l'ordre de 20 mm de rayon et de 10 mm d'épaisseur et non visible de la surface du fantôme (figure 13). Nous avons même recouvert le fantôme d'une feuille de papier blanc pour être sûrs que le modèle de tumeur soit invisible et pour pouvoir marquer les marges à l'aide d'un stylo.

L'utilisateur (dans notre cas un chirurgien) place la sonde échographique dans le fantôme, déline à la main les contours du modèle de tumeur dans l'image échographique, et, à l'aide des contours et marges de résection projetées dans le HMD (figure 14-a), marque avec un stylo les marges de la tumeur et les marges de résection sur le papier entourant le

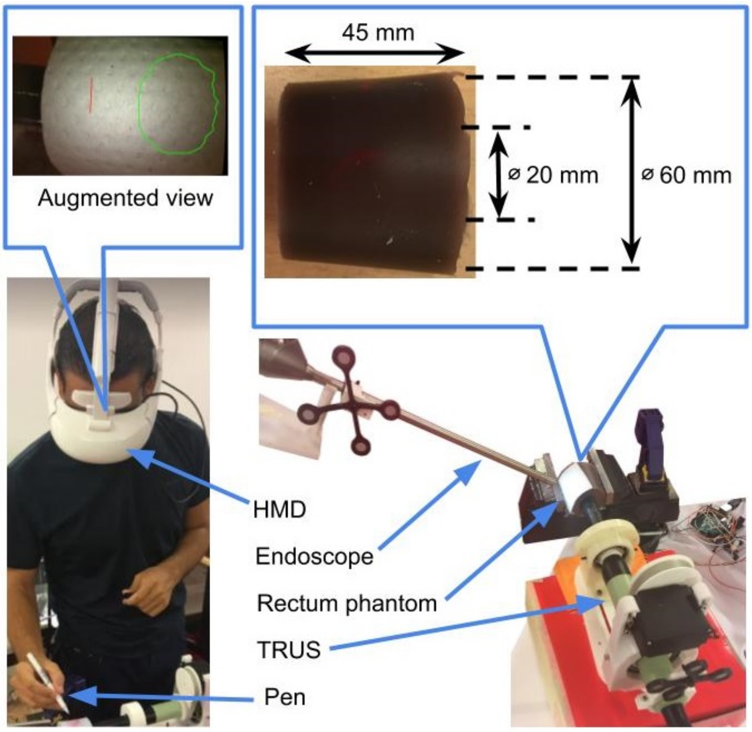


FIGURE 13 – Expérimentation sur le modèle physique de tumeur du rectum.

fantôme. Ceci simule la phase de marquage de la marge de résection sur le bas-rectum avec le bistouri électrique effectué classiquement en chirurgie robotisée.

Dans la procédure chirurgicale, seul le marquage de la marge de résection est utile, nous avons toutefois demandé à l'utilisateur de marquer les marges de la tumeur afin de vérifier la bonne superposition du modèle virtuel sur le modèle réel de tumeur.

Nous avons ensuite sectionné le cylindre aux endroits marqués par le chirurgien (figure 14-b). Nous constatons que le modèle de tumeur, teinté en rouge, affleure bien dans les marges de la tumeur définies en réalité augmentée. Par mesure nous constatons aussi que la marge de résection a bien été placée à 10 mm de la partie proximale de la tumeur.

L'échographie transrectale est utilisée de manière courante pour le diagnostique des tumeurs rectales mais à notre connaissance, cette expérimentation est la première preuve de concept de l'utilisation de l'échographie transrectale en chirurgie laparoscopique des tumeurs rectales.

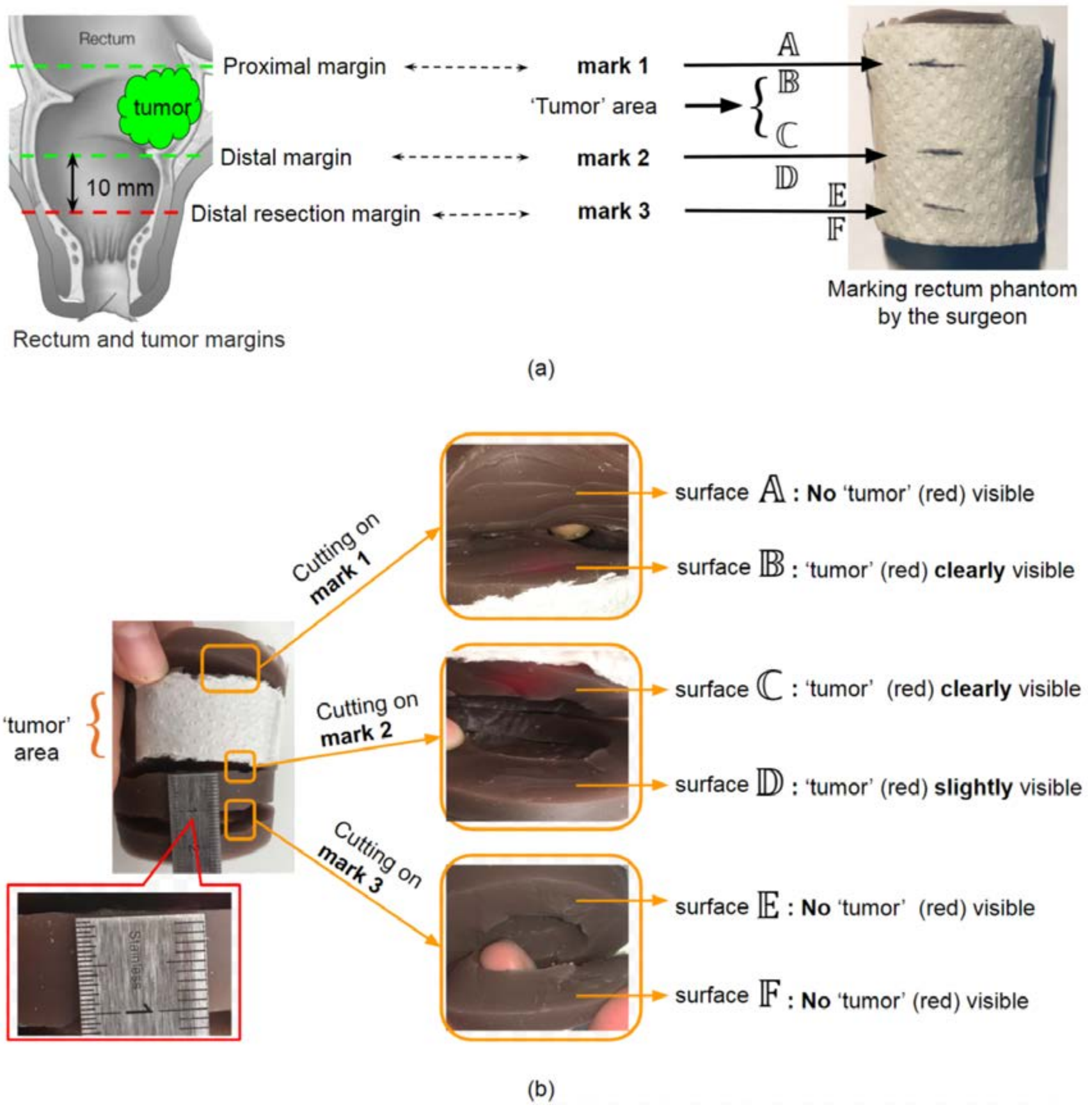


FIGURE 14 – Résultats de l'expérimentation sur le modèle physique de tumeur du rectum. a) vue augmentée de l'endoscope et résultats des marquages des marges ; b) plan de coupes selon les marges marquées.

Cette expérimentation nous a permis de démontrer la faisabilité de l'utilisation de la réalité augmentée dans un contexte proche de la clinique. Les méthodes de calibration sont suffisamment simples, automatiques et rapides (5 min pour l'échographe 3D, 10 min pour la calibration de l'endoscope) pour être utilisées en salle d'opération avant les interventions. Après discussion avec nos partenaires médicaux des résultats obtenus par notre expérimentation, nous envisageons les actions suivantes : l'intégration d'une vraie sonde transrectale 3D afin d'éviter les mouvements de rotation de notre sonde 3D maison ; la génération automatique ou semi-automatique de la marge de résection distale ; et l'utilisation de vrais outils chirurgicaux pour marquer la marge de résection.

6 Conclusion

Dans cette Thèse, nous avons développé une solution de réalité augmentée pour le guidage de gestes en chirurgie robotisée. Cette solution a été adaptée à deux activités cliniques bien définies : la chirurgie robotisée transorale des tumeurs de bases de langue et la chirurgie laparoscopique robotisée des tumeurs du bas rectum. Dans cette solution, la cible clinique (la tumeur) est acquise à l'aide d'un échographe 3D (en voie sous-mentonnière pour la tumeur de langue et en transrectal pour la tumeur du rectum), la tumeur est segmentée afin de créer un objet virtuel, cette information est alors projetée sur la vue endoscopique afin de permettre au chirurgien de localiser la position réelle de la tumeur et les marges de résection dans le champ opératoire. D'un point de vue méthodologique, nous avons imaginé et développé une solution originale de calibration des échographes 3D. Cette solution, utilisant à la fois un fantôme de calibration très simple (fabriqué en CAO) et un protocole allégé, a permis de confier cette calibration à des personnes non spécialisées et de raccourcir le temps nécessaire pour la calibration (5 min), la rendant disponible en conditions cliniques en salle d'opération. La calibration de l'endoscope 3D et la localisation 3D des systèmes d'acquisition ont été résolues de manière plus classique. Une première évaluation nous a permis de quantifier que les erreurs de localisation après reprojection sur les images endoscopiques étaient inférieures au millimètre.

Nous avons validé cette chaîne de réalité augmentée lors de trois expérimentations.

La première expérimentation, menée sur un fantôme en silicone avait pour objectif de comparer notre solution avec la solution habituellement utilisée par les chirurgiens pour localiser une tumeur (visualisation de la cible sur de l'image préopératoire -scanner X ou IRM- afin de se faire une image mentale de la pose de la tumeur et palpation de la zone avant le geste). Cette expérimentation a permis de prouver l'apport de la réalité augmentée.

Une seconde expérimentation avait pour objectif de vérifier l'applicabilité de la solution de réalité augmentée à la chirurgie transorale de la base de langue. Dans cette expérimentation, un modèle de tumeur en silicone avait été inséré dans la base d'une langue de mouton *ex-vivo*. Lors de cette expérimentation, grâce à la réalité augmentée, nous avons réussi à réséquer la tumeur avec la marge souhaitée de 10 mm avec une tolérance de ± 2 mm.

La troisième expérimentation avait pour objectif de vérifier l'applicabilité de la solution de réalité augmentée à la chirurgie laparoscopique des tumeurs du bas rectum. Dans cette chirurgie, la marge de résection entre la limite proximale de la tumeur et les sphincters du patient a une importance capitale pour sa vie après ablation. Nous avons simulé le rectum et la tumeur à l'aide d'un fantôme en silicone. Notre solution de réalité augmentée a permis de définir de manière très précise les limites proximales et distales de la tumeur (après incision du fantôme selon ces limites, la tumeur affleurait bien les plans de coupes) et la marge de résection était bien à la distance souhaitée de la limite proximale de la tumeur.

Les futurs travaux vont porter sur la segmentation automatique de la tumeur et la validation de la chaîne de réalité augmentée dans un vrai contexte de chirurgie robotisée à l'aide de la plateforme de recherche RAVEN.

List of Figures

Introduction

Robot-assisted surgery is a kind of minimally invasive surgery (MIS), in which an endoscopic camera and two dedicated surgical instruments are usually used to reach the operative field (mostly, internal organs) through small incisions or a natural orifice (*e.g.* mouth, nostril or anus). Such minimally invasive interventions offer patients a number of benefits: for instance, small incisions, low risk of infection and quick recovery. However, some drawbacks — such as limited field of view of the endoscopic camera, loss of tactile feedback from the surgical instruments and inability to perform intraoperative tissue diagnosis — make MIS more difficult for surgeons to perform than open surgery. Furthermore, it is challenging for surgeons to identify tumor boundaries accurately during MIS, due to the following reasons:

— Tumors are often partially or completely concealed in organs, so that tumor boundaries cannot been visualized with an endoscopic camera.

— Manual palpation is insufficient or impossible to perform during MIS. For instance, in transoral robotic surgery (TROS), manual palpation roughly shows to surgeons the superficial part of base of tongue tumors but not the tumor extent. Furthermore, in some surgeries, *e.g.* robot-assisted laparoscopic liver surgery, livers are not accessible for manual palpation. Currently, laparoscopic instruments or robotic instruments like da Vinci® surgical instruments have not provided tactile feedback.

— Preoperative imaging, such as magnetic resonance imaging (MRI) or computed tomography (CT), are usually performed several hours or several days before a surgery, so that those preoperative images cannot reflect accurately the locations of tumors in deformed operative fields. Intraoperative tissue deformation often happens, because of patient repositioning, organ displacement or surgical manipulation (*e.g.*

> insufflation of gas into the abdominal cavity to create the pneumoperitoneum for laparoscopic surgery). Therefore, surgeons use intraoperative images for surgical guidance. However, common intraoperative imaging techniques have a variety of limitations: For example, open MRI and cone beam computed tomography (CBCT) (which are used in high-cost hybrid operating rooms) provide poor-quality images of soft tissues; fluoroscopy carries the risk of radiation; intraoperative ultrasonography (US) images are difficult to understand, thereby costing surgeons' time to identify tumor boundaries.

In clinical practice, during image-guided surgery, surgeons mentally deform and map patient-specific medical images to operative fields to guide tumor resection. Such mental practices is highly dependent on surgeons' knowledge of anatomy and experience, that may cause compromised resection margins. Incomplete resection of soft-tissue tumors shows the high risk of cancer recurrence [Vermaas et al., 2007], so it is particularly important to provide solutions for accurate tumor resection.

Augmented reality (AR) technique has been used in many surgical procedures, such as neurosurgery [Masutani et al., 1998], breast biopsy [Rosenthal et al., 2002] and laparoscopic liver surgery [Feuerstein et al., 2008], for intraoperative guidance. It generates a tumor model from medical images, then, registers and superimposes the tumor model onto surgical scenes, thereby showing to the surgeon the tumor area in the operative field. Preoperative MRI and CT images usually have high resolution and good quality, and are therefore ideal for use in clinical AR applications to create tumor models. In order to take intraoperative tissue deformation into account, a number of studies propose to register the preoperative images to surgical scenes via intraoperative CBCT or fluoroscopy images [Mountney et al., 2014; Liu et al., 2015]. However, volume-based deformable registration (*e.g.* MRI/CT to CBCT [Reaungamornrat et al., 2013], or MRI/CT to US [Lange et al., 2003; Hu et al., 2012] has been a challenging task. An alternative could be to use intraoperative images to generate tumor models, thereby avoiding the complicated cross-modality deformable registration. Intraoperative US is superior to other intraoperative imaging techniques in the context of soft tissue surgery, because US is ideally suited for imaging soft-tissue structures which are not well imaged by X-rays-based CBCT and fluoroscopy. Moreover, US imaging technique does not carry the risk of radiation-induced injuries and has no known harmful effects on patients and medical staffs. Another intraoperative imaging modality that is suited for diagnosing soft tissue tumors is Open MRI, but this imaging technique shortens MRI scanning at the expense of imaging resolution. Furthermore, compared to US imaging systems which are widely available in hospitals, Open MRI systems are used only in high-cost hybrid operating rooms.

Indeed, as early as 1979, Dr. Bernard Sigel has performed US imaging on patients for intraoperative visualization of biliary calculi [Sigel et al., 1979]. Since then, this imaging technique has been used for a variety of purposes, such as non-palpable breast masses visualization [Schwartz et al., 1988], tissue shift correction in neurosurgery [Comeau et al., 2000], surgical navigation in liver surgery [Kruskal and Kane, 2006]. Recent advances in US system design and manufacture — it includes the improvement in digital systems and US image resolution, the development of three-dimensional (3D) and four-dimensional (4D, three-spatial dimensions plus one-time dimension) transducers, and the improvement in portability of US machines — greatly enhance the US imaging performance. For instance, 2D array US transducers have been proposed to provide volumetric imaging, thereby producing 3D US images in real-time [Yen et al., 2000]. Moreover, a number of manufacturers, such as Clarius Mobile Health, Siemens Healthineers, Philips, Sonoscanner and so on, have developed advanced portable US systems, that significantly facilitates the intraoperative use of US imaging technique.

Based on the intraoperative US imaging technique, this thesis is to implement an AR framework for tumor visualization and resection guidance. The framework is proposed with clinical partners, in order to adapt it to two specific surgical procedures:

- Transoral robotic surgery for base of tongue cancer operated by the otolaryngology, head and neck surgery department of the Montpellier hospital (*Centre Hospitalier Universitaire de Montpellier, France*)
- Robot-assisted laparoscopic surgery for low-rectal cancer operated by the colorectal surgery department of Montpellier cancer institute (*Institut du Cancer de Montpellier, Val d'Aurelle, France*)

In these two surgical procedures, surgeons face the same intraoperative challenge — tumors are concealed in the base of tongues (or the rectal walls) and cannot be visualized with endoscopic cameras. The proposed framework overcomes this challenge by following steps: During these two surgical procedures, the tumor can be imaged by an US system to create virtual tumor model, then, the tumor model is superimposed onto the 3D endoscopic view of an operative field according to the spatial registration of the US images to the endoscopic camera.

The work presented in this thesis is based on the context of robot-assisted minimally invasive surgery. It was co-funded by Labex CAMI[4] and the Bretagne region (France). This research work was advised by Nabil ZEMITI (University of Montpellier) and supervised by Jean-Louis DILLENSEGER (University of Rennes 1) and Philippe POIGNET (University of Montpellier). It was conducted in the IMPACT team of the Laboratoire Traitement du Signal et de l'Image (LTSI, Rennes) and the DEXTER team of the Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (LIRMM).

Organization of the dissertation This dissertation shows the medical context, state-of-the-art methods of AR guidance in minimally invasive surgery and our studies of implementing an intraoperative US-based AR framework for guiding tumor resection in robot-assisted base of tongue cancer and low-rectal cancer. The organization of this thesis is as follows:

Chapter 1 describes the medical context of the thesis. It introduces robot-assisted surgical procedures and current intraoperative image-guidance systems for the base of tongue cancer and the low-rectal cancer treatments. Moreover, the performance of US systems showing base of tongue tumors and low-rectal tumors is introduced in this chapter too.

Chapter 2 presents the state-of-the-art methods of implementing surgical AR systems. Furthermore, current US-based AR systems and their limitations in minimally invasive interventions are also summarized in this chapter.

Chapter 3 presents the main contributions of this thesis: the implementation and the evaluation of the proposed intraoperative US-based AR framework, according to the surgical workflow of (1) transoral robotic surgery for base of tongue cancer and (2) robot-assisted laparoscopic surgery for low-rectal cancer. A fast and accurate 3D US probe calibration is proposed to facilitate the intraoperative use of US imaging technique.

Chapter 4 describes experimental protocols which simulate the integration of the implemented framework into the surgical workflow. A silicone rectum phantom and an ex-vivo lamb tongue phantom were used in these experiments. The experimental re-

---

[4] Computer Assisted Medical Interventions: http://cami-labex.fr/

sults show that, according to the augmented endoscopic views, a surgeon is able to accurately localize the boundaries of the simulated tumors in these phantoms.

Finally, the dissertation is concluded by summarizing the proposed study and discussing some future work.

Chapter 1

Medical context and image-guidance systems

Contents

This chapter introduces two surgical procedures: Transoral robotic surgery for base of tongue cancer and robot-assisted laparoscopic surgery for low-rectal cancer. In these two surgical procedures, surgeons face the same intraoperative challenge, which is the difficulty to localize base of tongue tumors or low-rectal tumors in the intraoperative endoscopic views. However, this challenge can be overcame by the same technique. It uses intraoperative US to image an operative field and target the tumor, then, augments the endoscopic views with tumors information from US images via AR technique. Furthermore, this chapter presents current studies in developing image-guidance systems based on above medical context. In addition, the performance of US imaging technique showing base of tongue tumors and low-rectal tumors are introduced.

1.1 Transoral robotic surgery for base of tongue cancer

1.1.1 Surgical treatment for base of tongue cancer

In anatomy, the tongue is divided into an anterior part (the oral tongue) and a posterior part (the base of the tongue). As illustrated in Figure 1.1, the base of tongue is located behind the circumvallate papilla of the tongue and is a part of the oropharynx. The base of tongue is intricately involved in people's respiration, articulation and deglutition. Cancer developed in the base of tongue is a type of oropharyngeal cancer. In 2018, there are more than 92,800 new cases of oropharyngeal cancer and over 51,000 associated deaths being estimated all over the world [Bray et al., 2018]. Furthermore, base of tongue cancer affects directly patients' breath, speech and swallow functions. Surgery is an effective treatment for the most of base of tongue cancer, that is primarily dictated by the histopathologic diagnosis and tumor size. In addition, radiation therapy and chemotherapy are used to kill remained cancerous cells.

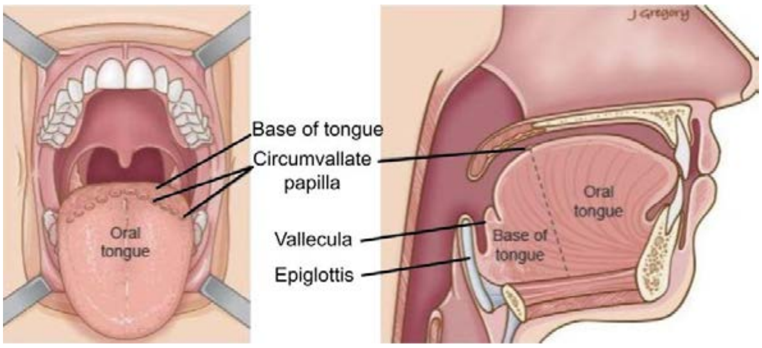


Figure 1.1 – The base of tongue. Source: Head & Neck cancer guide® tongue base cancer.

The base of tongue is a hard-to-reach area and some surgeries require suprahyoid incision to access to this area. With the advent of improved surgical robots and robotic surgical instruments, surgeons reach the base of tongue area through patients' mouth. The transoral approach reduces the risk of infection and shortens patients' recovery time. Transoral robotic surgery is ideal for treating the benign tumor with limited extension into surrounding tissues, T1 and T2 malignant tumors, or exophytic T3 malignant disease [Van Abel and Moore, 2013]. In an operating room, a patient is orally or nasally intubated with an endotracheal tube. Next, a surgeon palpates the patient's base of tongue with his/her finger, thereby reassessing the tumor information including its location, extent and mobility. Meanwhile, the surgeon formulates a mental image of the tumor information. Then, the surgeon places an oral retractor on the patient's tongue to provide adequate exposure of the base of tongue, as shown in Figure 1.2. Finally, the surgeon controls a robot such as the da Vinci® surgical system to remove the tumor from the base of tongue (Figure 1.2 and Figure 1.3). Throughout the resection, the surgeon should provide an adequate pathologic margin of resection.

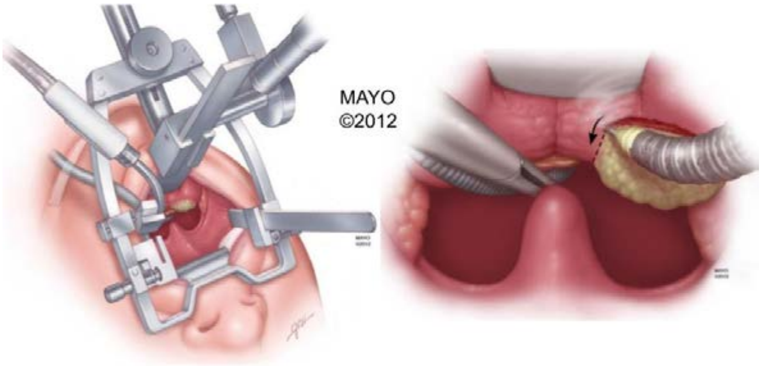


Figure 1.2 – Transoral robotic surgery for the base of tongue cancer. Source: [Van Abel and Moore, 2013].

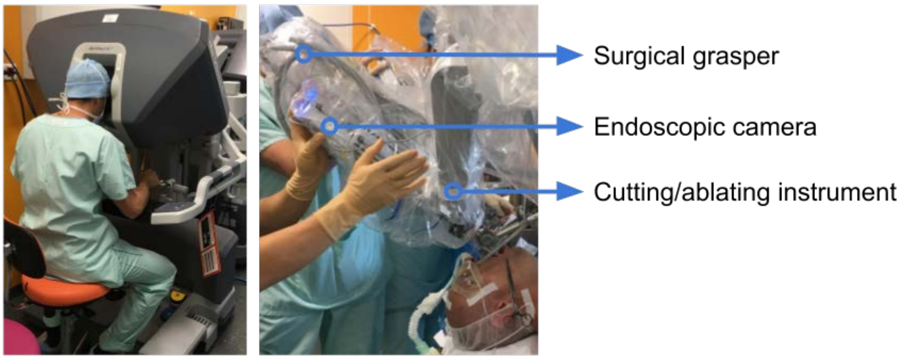


Figure 1.3 – The console (left) and three robotic arms respectively holding a surgical grasper, a cutting/ablating instrument and a 3D endoscopic camera (right) of the da Vinci® surgical system used in transoral robotic surgery.

It is challenging for surgeons to perform transoral robotic surgery on patients' base of tongue, due to following reasons: First, tumors are often concealed in tongues, and surgeons cannot see the whole tumor boundaries; second, it is difficult for surgeons to tell the difference between cutting pathologic tissues versus cutting normal tissues through robotic surgical instruments; third, based on the manual palpation at the beginning of surgery, surgeons have to formulate the map of tumor location, extent and mobility in their mind, and use this information to guide the tumor resection. This practice is highly dependent on surgeons' experience and requires a difficult initial learning process. Many studies have proposed image-guidance systems to overcome above difficulties, as presented in the following section.

1.1.2 Image-guidance systems in transoral robotic surgery

A number of clinical studies in the field of image-guided transoral robotic surgery for oropharyngeal cancer have been proposed. For instance, Desai *et al.* has demonstrated a CT image-guidance system in the da Vinci® robotic surgery, where the CT image and a pointer are registered to a patient and the trajectory of the tip of the pointer is displayed on the CT image on an external screen [Desai et al., 2008]. The main concern of this preoperative CT image-based approach is that the spatial registration between the CT image and the patient is based on some bony landmarks. However, the manipulation of the patient — the tongue is retracted as shown in Figure 1.2 — may change the relationship between the tumor and the bony landmarks, thereby leading to inaccurate intraoperative guidance. A recent study proposed by Clayburgh *et al.* used intraoperative transoral US technique, prior to starting the tumor resection, to show the tumor edge and important vessels to a surgeon [Clayburgh et al., 2016]. However, the surgeon needs to extract tumor information from a group of 2D US images and mentally transform the information to the operative field. The output quality of this mental work highly depends on the surgeon's training, experience and his/her thorough knowledge of human anatomy on US images.

Rather than displaying guidance images on an external screen, it is advantageous to provide an augmented visualization which fuses the surgical scene with the information extracted from the medical images. To this end, some feasibility studies have been proposed. Pratt *et al.* augmented intraoperative endoscopic views with mesh models derived from preoperative CT and MRI images, but the image-to-video registration does not take the intraoperative tongue deformation into account [Pratt and Arora, 2018]. To account for the gross deformation (> 30 mm) due to the retraction of a tongue, Liu *et al.* and Reaungamornrat *et al.* developed an image guidance system including: First, deformable registration between the intraoperative cone beam computed tomography (CBCT) and preoperative MRI (or CT) of a tongue [Reaungamornrat et al., 2013]; and second, CBCT-to-video registration based on some artificial markers which are attached to the tongue surface [Liu et al., 2013, 2015]. But the overall runtime of such a system is quite slow and the step of deformable image registration already takes approximately 5 minutes [Reaungamornrat et al., 2013]. Furthermore, these artificial markers are interruptive for surgical workflow.

Recently, some studies have demonstrated an intraoperative fluorescence spectroscopy-based guidance system in transoral robotic surgery [Patsias et al., 2015; Farwell et al., 2016]. However, its performance in localizing deep margins of tumors in tongues is not satisfactory. We believe that using 3D US imaging technique will improve the intraoperative guidance, because it is easier to delineate soft-tissue tumors in 3D US images than in fluoroscopy images. Moreover, contrary to fluoroscopy, US imaging technique does not carry the risk of radiation-induced damage for surgeons and patients.

1.1.3 Ultrasonography of base of tongue cancer

Conventionally, MRI, CT or positron emission tomography (PET) images are used for base of tongue cancer management. However, these imaging techniques have disadvantages such as missing primary tumor in PET images [Branstetter et al., 2005], dental artifact in CT images [Trotta et al., 2011], weak identification between small tumor and surrounding lingual lymphoid tissues in MRI images [Trotta et al., 2011]. Alternatively, recent studies present that US imaging systems are comparable and complementary to CT and MRI systems for the base of tongue cancer evaluation [Blanco et al., 2014; Coquia et al., 2015]. Moreover, Blanco *et al.* performed US imaging on the submental area of 22 patients and found that 100% clinically suspicious base of tongue tumors, tumor extent and their relationship with surrounding tissues can be visualized in US images, where the smallest tumor is between 5.4 mm and 10.4 mm in each of three dimensions [Blanco et al., 2014]. Based on the study of Blanco *et al.*, the same team investigated and presented the potential of using US systems to localize primary base of tongue tumors and other important oropharynx area [Fakhry et al., 2014].

To date, there are only few studies using US systems in transoral robotic surgery. This may be due that it is difficult to map the tumor information from US images to surgical scenes (intraoperative endoscopic views). One recent study proposed by Clayburgh *et al.* used intraoperative US images to show base of tongue tumors and other oropharyngeal tumors on an external screen; then, a surgeon extract tumor information from US images and mentally transform the information to the intraoperative endoscopic views of the operative field[Clayburgh et al., 2016]. The surgical outcomes show that intraoperative US images are helpful in guiding tumor resection. Moreover, the measurements of the tumor via US are the same as the measurements of pathological specimens [Clayburgh et al., 2016].

The previous cited studies show the ability of US imaging techniques to localize base of tongue tumors and present the potential of intraoperative US-based AR application in transoral robotic surgery. However, to the best of our knowledge, the intraoperative US-based AR guidance technique has not been integrated into the workflow of transoral robotic surgery.

1.2 Robot-assisted laparoscopic low-rectal cancer surgery

The rectum starts as a continuation of the sigmoid colon and ends by becoming a continuation of the anal canal, as shown in Figure 1.4. It serves as a temporary reservoir for feces storage, and plays an important role in controlling defecation. Rectal cancer develops from the lining of rectums. Worldwide, colorectal cancer is one of the most commonly diagnosed cancer and the second leading cause of cancer death, wherein more than $704,000$ new cases of rectal cancer and over $310,000$ associated deaths are predicted in 2018 [Bray et al., 2018].

1.2.1 Surgical treatment for low-rectal cancer

Surgery is the main treatment for rectal cancer. Surgical robotic systems, such as da Vinci® surgical platforms, facilitate the accessing to patients' narrow pelvis and improve surgeons' gestures accuracy [Baek et al., 2015]. For example, during robot-assisted laparoscopic surgery, as shown in Figure 1.5, a patient is positioned on the patient cart of a da

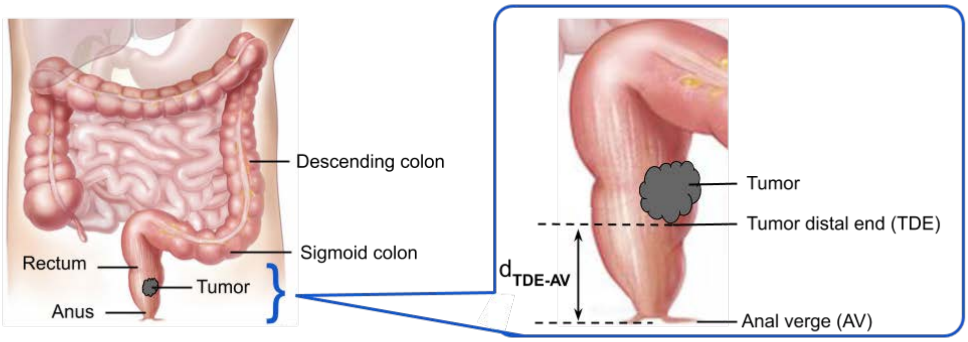


Figure 1.4 – The location of a rectal tumor. Figure reproduced from Digestive Appareil Virgin.svg (author: William Crochot), used under Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).

Vinci® system, where robotic arms hold surgical instruments and an endoscope to reach the rectum area through the patient's abdomen. These robotic arms are controlled by a surgeon from the da Vinci® console.

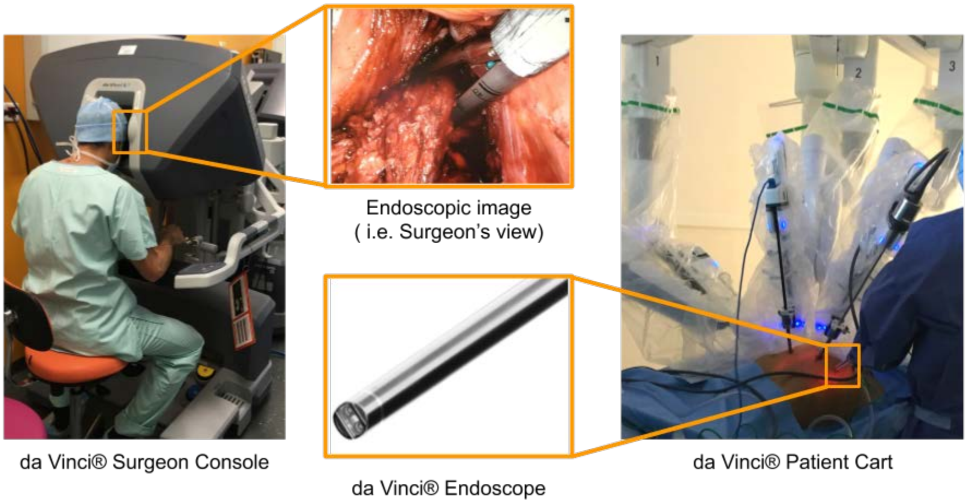


Figure 1.5 – Robot-assisted laparoscopic colorectal surgery using a da Vinci® surgical platform.

Conventionally, a tumor having its distal end within 60 mm of the anal verge is a low-rectal tumor ($d_{\text{TDE-AV}} < 60$ mm in Figure 1.4), otherwise it is a high or mid-rectal tumor [Rullier et al., 2013]. For treating high and mid-rectal cancer, the surgical procedure is well standardized: partial or total mesorectal resection and sphincter preservation [Rullier et al., 2013]. It is important to preserve patients' anal sphincters, because the function of anal sphincters is to maintain fecal continence. However, the procedure of sphincter preservation for a patient with low-rectal cancer is very complicated, and requires different surgical techniques according to how close the low-rectal tumor is to the anal sphincters. If the distal end of a low-rectal tumor is more than 10 mm from the anorectal ring, as shown in Figure 1.6, this type of tumor is called supra-anal tumor [Rullier et al., 2013]. One oncological rule is to perform distal resection 10 mm below the distal ends of supra-anal tumors, thereby preserving anal sphincters. The surgical procedure for supra-anal tumor resection by a da Vinci® system is shown in Figure 1.7 and major steps are presented as follows:

— **Distal resection**: a surgeon defines the distal resection margin approximately 10 mm below the distal end of the tumor according to the result of digital rectal examination (*i.e.* manually palpating the rectum through the anus to show tumor area). Afterward, the surgeon performs distal resection on the rectum with a da Vinci® stapler.

— **Colon resection**: the surgeon takes the colon out of the patient's abdomen and cuts it in the middle of the descending colon, which is around the inferior mesenteric artery.

— **Colorectal anastomosis**: the surgeon performs ultra-low colorectal anastomosis with double-stapled technique to connect the remaining colon to the remaining rectum.

In this procedure, accurately defining the distal resection margin is one of the most challenging tasks for surgeons, because supra-anal tumors are not visible in intraoperative endoscopic views. Moreover, digital rectal examination is not sufficient to identify clear tumor boundaries.

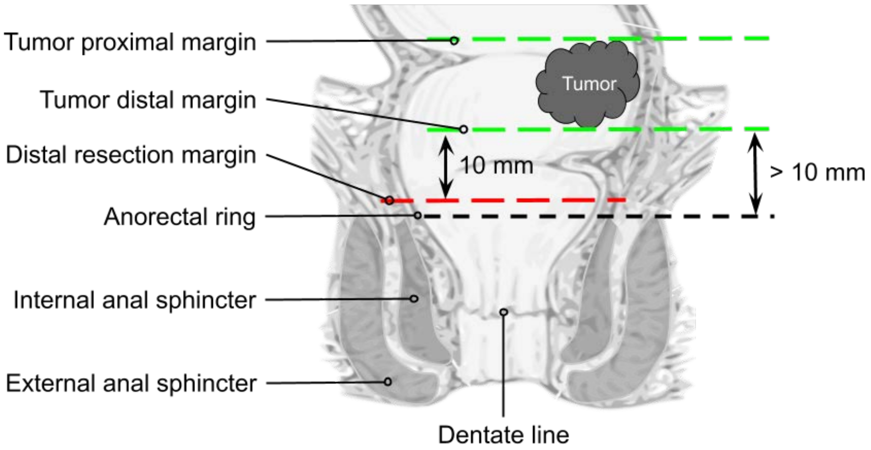
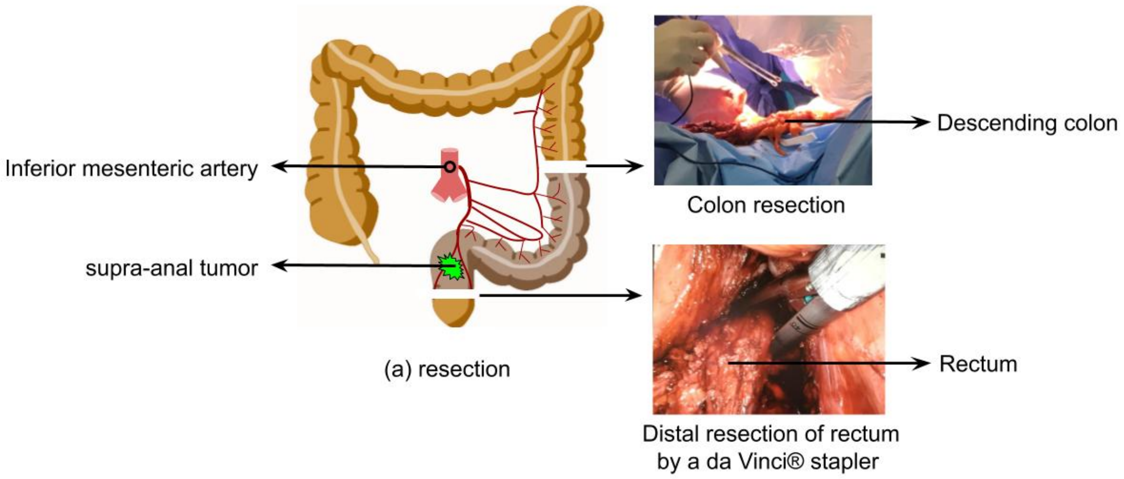
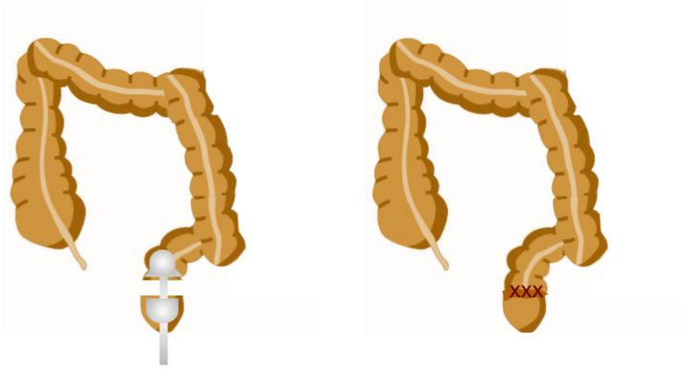


Figure 1.6 – The location of a supra-anal tumor in a rectum. Figure reproduced from Anorectum-en.svg (author: National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)), used under Attribution-ShareAlike 4.0 International (CC BY-SA 4.0).



(a) resection



(b) Colorectal anastomosis by double-stapled technique

Figure 1.7 – Resection of supra-anal tumor and ultra-low colorectal anastomosis.

1.2.2 Image-guidance systems in robot-assisted rectal cancer surgery

To the best of our knowledge, to date, only one technique has been proposed: In June 2016, a Florida Hospital team demostrated a preoperative MRI-based navigation system in transanal endoscopic microsurgery [Atallah et al., 2016]. They placed several markers on the surface of a patient's abdomen before MRI imaging, and kept the markers in the same position on the patient until the end of the surgery. By tracking these markers during surgery, the tip of a tracked surgical tool and the patient's MRI image are registered to the same coordinate system — the patient's abdomen surface, thereby showing on the MRI image the trajectory and position of the tip of the surgical tool. As a result, however, this marker-based registration procedure does not take tissue deformation beneath the abdomen surface into account properly. Hence, intraoperative imaging techniques provide an alternate solution:

Indeed, intraoperative US allows surgeons to adequately track rectal tissue deformation, thereby guiding rectal tumor resection accurately. The study in [Greif et al., 2009] displayed intraoperative US images on an external screen, in order to guide the laparoscopic colorectal resection for 25 patients including 6 cases of rectal lesions. However, during the surgery, the surgeon needs to mentally transform an US image in a good position, orientation and scale, in order to map it on the surgical scene. As a result, the positive surgical outcomes highly depend on surgeons' experience and knowledge of anatomy in US images. Such mental load of surgeons can be reduced by an AR guidance system which superimposes the tumor information on the surgical scene.

1.2.3 Ultrasonography of rectal cancer

Current available methods for rectal tumor assessment include CT, MRI, US and a manual palpation test (*i.e.* digital rectal examination). The digital rectal examination is the most simple and common method, but it allows only to localize roughly the tumor area. This is due to that the palatable interface between the lesion and the surrounding normal tissues is usually subtle. Moreover, it is difficult to recognize small lesions in rectums with digital rectal examination. A comparative study of transrectal US, pelvic CT and MRI in rectal cancer staging was proposed in [Kim et al., 1999]: They measured the depth of invasion of rectal tumors for 89 patients by transrectal US, 69 patients by pelvic CT and 73 patients by MRI. As a result, the overall accuracy of cancer staging was 81.1% (79 out of 89 patients) by transrectal US, 81% (59 out 73 patients) by MRI and 65.2% (45 out of 69 patients) by CT. The comparative study shows US is superior to MRI and CT in rectal tumor staging. Furthermore, surgeons from the study of [Greif et al., 2009] used intraoperative US images to guide rectal tumor resection for 6 patients, and reported that US imaging technique is an effective sole method for early rectal cancer localization. While US images are able to show rectal tumors, integrating the US segments of tumors into surgical scenes via AR techniques would provide further benefits, that has not been proposed in any other studies.

1.3 Conclusion

The previous cited studies show that US images have been widely used for base of tongue cancer and rectal cancer staging. Moreover, compared to other intraoperative imaging techniques like CBCT or fluorescence, US imaging technique does not carry radiation-induced injuries for patients and surgeons. Furthermore, US systems are much cheaper to use than CBCT and fluorescence in operating rooms. These advantages of US imaging technique motivate our proposition of an intraoperative US-based AR framework for guiding resection of base of tongue tumors or low-rectal tumors in robot-assisted surgery, where tumor information from intraoperative images is superimposed onto intraoperative endoscopic views.

Chapter 2

Augmented reality and intraoperative US for surgical guidance

Contents

This chapter presents the state-of-the-art of augmented reality (AR) techniques in minimally invasive surgery including the patient-specific virtual model generation, the visualization systems, and the registration methods to align the virtual models to the surgical scenes. Then, studies about using intraoperative ultrasonography (US) in surgical AR systems to guide soft tissue surgery are introduced. Finally, the problems of using AR and intraoperative US technique to implement a framework for tumor resection guidance in our specific surgical procedures (*i.e.* robot-assisted surgery for treating base of tongue cancer and low-rectal cancer) are described.

2.1 Introduction

AR is a technology of superimposing virtual objects on the real-world environment in real-time. It is different from virtual reality which creates its own environment by stimulating a person's vision and hearing via computer. Azuma *et al.* defines an AR system as a system with following properties [Azuma, 1997]:

— Real-world and virtual information are fused and coexisted in users' view.

— The system runs in real-time, *i.e.* real-world and virtual information are interactive in real-time.

— Real-world and virtual information are registered to each other in three dimensions (3D).

In minimally invasive surgery, a 2D or 3D endoscopic camera is used to visualize operative fields. Figure 2.1 shows the procedure of augmenting an endoscopic view (*i.e.* surgeon's view):

(1) Generating a virtual object (*e.g.* a 3D rendering of tumor surface) from preoperative or intraoperative medical images, tracked surgical instruments, or preoperative plans (*e.g.* resection margins) (details in section 2.2.1).

(2) Capturing an operative field in real-time via an endoscopic camera.

(3) Registering the virtual object to the endoscopic images in real-time via an AR system (details in section 2.2.3).

(4) Superimposing the virtual object on the endoscopic images in real-time via the AR system, and displaying these augmented images to a surgeon (details in section 2.2.2).

AR systems facilitate surgical procedures and improve surgical outcomes by providing visualization aid for surgeons. For example, AR systems can give to surgeons 'X-ray vision' of tumors and vessels — these anatomical structures are usually concealed in organs and neither visible with endoscopic camera nor palpable during minimally invasive interventions. Such visualization aid improves the accuracy of tumor resection and avoids perforating unsuspected vessels [Liu et al., 2013; Mountney et al., 2014]. In some image-guided surgeries, surgeons use intraoperative US or other intraoperative imaging techniques to identify tumors, then, mentally transform the tumor information into the operative fields [Greif et al., 2009; Clayburgh et al., 2016]. In this case, AR systems greatly facilitate these image-guided

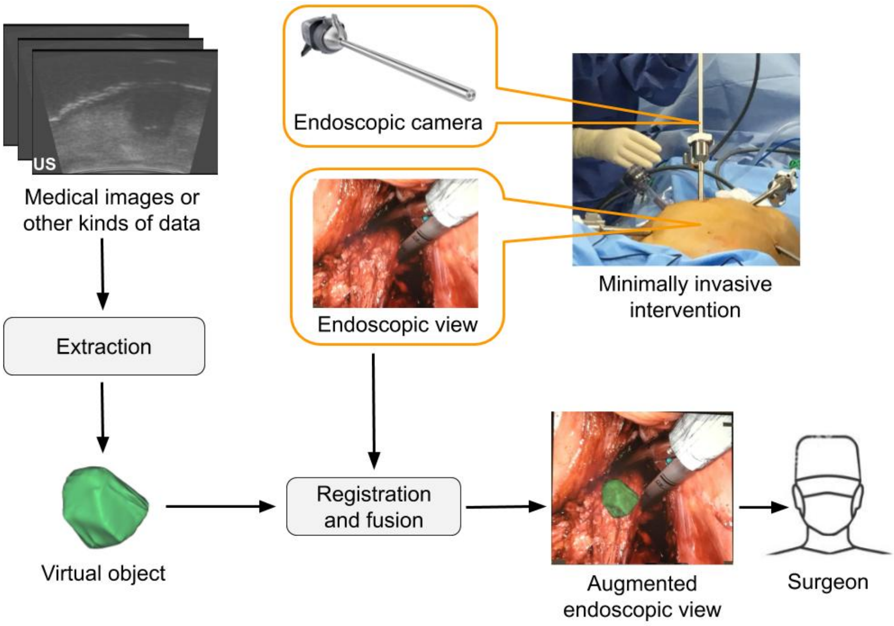


Figure 2.1 – The procedure of implementing an AR system in minimally invasive surgery.

surgical procedures, decrease intervention time, and reduce surgeons' cognitive load by presenting tumors on surgeons' view.

Visualization systems used in minimally invasive surgery have limited field of view and decreased depth perception (mostly in laparoscopic surgery). These limitations can be overcame by AR techniques. For example, to improve the accuracy of tumor resection in partial nephrectomy, a study propose to show a surgeon the resection path underneath the tumor via AR techniques, and the experimental results prove that their AR system significantly improve the resection accuracy [Singla et al., 2017]. Moreover, the study proposed in [Haouchine et al., 2013] superimposes a liver biomechanical model onto a human liver undergoing deformation due to surgical instrument interaction during laparoscopic hepatic surgery, thereby helping surgeons to localize tumors in deformed livers.

In following sections, the state-of-the-art of implementing surgical AR systems are introduced, and the studies about intraoperative US-based AR systems are described.

2.2 Implementing augmented reality for surgical guidance

In order to implement AR systems into operating rooms for surgical guidance, following three challenges need to overcome:

(1) generating a patient-specific virtual model of surgical targets, as discussed in section 2.2.1.

(2) finding a way to display the augmented view on a visualization system, as presented in section 2.2.2.

(3) achieving and maintaining a high AR accuracy during surgery, as described in section 2.2.3.

2.2.1 Generating virtual objects

Medical images are often used to visualize internal structures of the body in order to diagnose, monitor or treat medical conditions. For instance, a surgeon use a patient's magnetic resonance imaging (MRI) data to diagnosing the patient's cancer and make surgical plans for this patient. In order to provide resection guidance in soft tissue surgery, a number of studies have proposed to extract the information of tumors, organs or vessels from patients' medical images to create patient-specific virtual objects for surgical AR systems [Haouchine et al., 2013; Mountney et al., 2014]. Commonly, these patient-specific virtual objects can be generated from patients' preoperative images (*e.g.* MRI or computed tomography (CT)) or intraoperative images like ultrasonography (US).

— **Generating virtual objects from preoperative images.** Preoperative MRI or CT images are produced by diagnostic imaging techniques several hours or several days before surgical interventions. These preoperative images show adequate information of internal organs, tumors or vessels. However, during soft tissue surgery, these preoperative images cannot reflect accurately the tumor locations and tumor boundaries in deformed operative fields. For example, as shown in Figure 2.2, in robotic laparoscopic liver surgery, the liver shape (including the tumor location and shape) after insufflation of abdominal cavity is significantly different from that in preoperative CT images, so that transforming the liver tumor model from the CT images onto the surgical scenes (*i.e.* the deformed liver) requires a complex deformation of the tumor model [Mountney et al., 2014]. This complex deformation is based on the registration between preoperative images and surgical scenes. In minimally invasive interventions, surgical scenes are provided by endoscopic cameras. As the endoscopic images shows only the surface information of operative fields and miss the information of the deformed tissues beneath the surface, registering the preoperative image to the endoscopic image is significantly difficult, as explained in detail in section 2.2.3. Therefore, finding a proper deformation for a virtual object which is generated from preoperative images remains a challenging task.

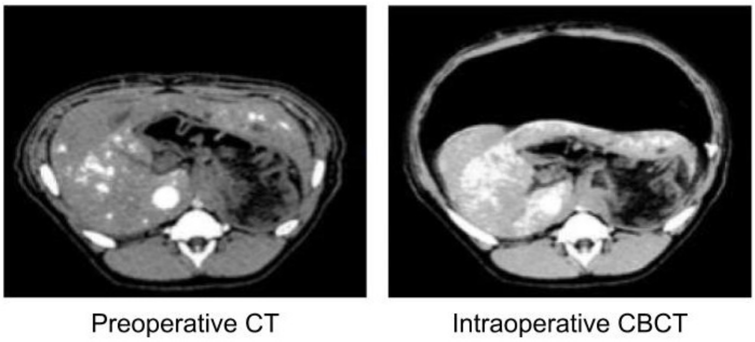
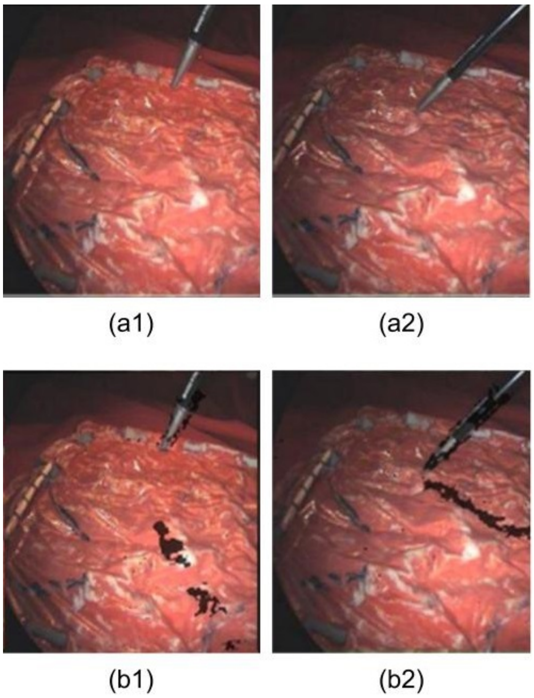


Figure 2.2 – Preoperative CT image and intraoperative CBCT image of the same patient's abdomen showing the significant deformation of the patient's abdominal organs. Source: [Mountney et al., 2014].

- **Generating virtual objects from intraoperative images.** Intraoperative images (such as cone beam computed tomography (CBCT) images shown in Figure 2.2) are often produced in surgery and show the lesion locations in operative fields properly. Therefore, the transformation of the tumor models (generated from the intraoperative images) onto surgical scenes does not require significant deformation of the tumor models. A number of studies have proposed to use CBCT and fluoroscopy in AR systems [Liu et al., 2013; Kenngott et al., 2014; Mountney et al., 2014]. However, these X-rays-based imaging modalities usually do not provide good imaging resolution of soft tissues and these images are difficult to be used to create tumor models. In the study presented in [Kenngott et al., 2014], a commercial AR system is used to register a patient's MRI image to the patient's CBCT image, then, register the CBCT image onto real-time fluoroscopy, thereby generating a real-time overlay of a liver-tumor model on the fluoroscopy for AR guidance. Indeed, this liver-tumor model is created from the patient's preoperative MRI image. An alternative could be to use intraoperative US image to generate the tumor models, since US imaging technique has been widely used to image soft-tissue tumors and cancer staging [Greif et al., 2009; Blanco et al., 2014; Kisansa and Andronikou, 2017]. For example, a study has proposed to create a prostate model from an intraoperative US image to augment the endoscopic view in laparoscopic radical prostatectomy [Simpfendörfer et al., 2011]. In the study, the prostate model was registered to the endoscopic image based on corresponding artificial fiducials in the US image and the endoscopic image. This is one of the main advantages of using intraoperative US images to create soft-tissue tumor models — avoiding the complex deformation of virtual models. Furthermore, many studies have proposed a variety of registration methods for merging intraoperative US images to endoscopic images, as presented in section 2.3.

In addition to generating patient-specific virtual objects from patients' medical images, some intraoperative tracking data can be used to augment the intraoperative endoscopic views. For example, in order to improve surgeons' depth perception in laparoscopic surgery, a study proposed in [Nicolaou et al., 2005] uses the tracking data of a laparoscopic instrument to create a virtual shadow of the instrument, then, superimposes the virtual shadow onto the laparoscopic view. As shown in Figure 2.3, when a laparoscopic instrument approaches to the surface of a phantom, it is difficult to identify how close between the instrument tip and the surface of the phantom (Figure 2.3 (a1) and (a2)), but this problem can be solved by the virtual shadow of the instrument displayed on the laparoscopic view (Figure 2.3 (b1) and (b2)). Moreover, some studies propose to create virtual objects from preoperative surgical plans, such as the safe path to reach the operative target [Liu et al., 2018; Gavaghan et al., 2012]. For instance, Figure 2.4 shows that the preoperative planning data (*i.e.* the location, direction and depth of needle insertion) can be projected onto a surgeon's view during a biopsy procedure [Gavaghan et al., 2012]. The augmented view guides the surgeon to insert the needle into a precise location (the left image of Figure 2.4) and shows to the surgeon the depth to reach the target (the right image of Figure 2.4).

Conclusion. In soft tissue surgery, a patient-specific tumor model for an AR system can be generated from the patient's medical images by image segmentation and surface modelling. In the work presented in this thesis, 3D US imaging technique was chosen. It is because, first, a number of studies have shown that base of tongue tumors and low-rectal tumors can be visualized with US systems, as described in section 1.1.3 and 1.2.3, so that tumors can be segmented from US images. Second, unlike preoperative MRI or CT images, intraoperative US images can properly show the tumor location in an operative field, thereby avoiding the complex deformation of the tumor model when transforming it onto the surgical scene. In addition to tumor models, tracking data of surgical instruments and preoperative surgical plans can be used to augment surgeons' views too.


(a1) (a2)

(b1) (b2)

Figure 2.3 – Augmenting the raw laparoscopic images (a1) and (a2) with the shadow of the laparoscopic instrument (b1) and (b2), when the instrument tip approaches to the phantom surface. Source: [Nicolaou et al., 2005].

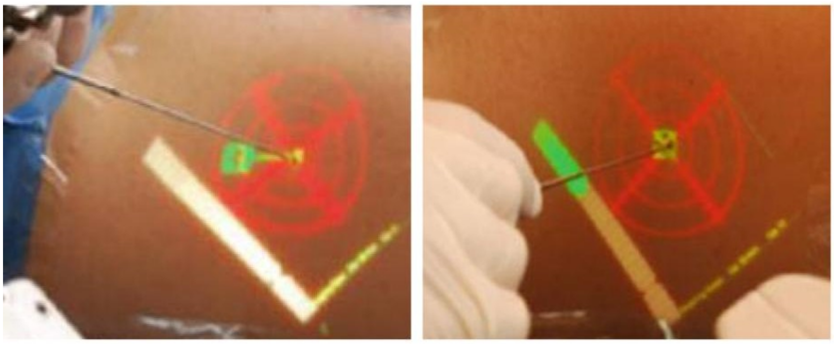


Figure 2.4 – Augmenting surgeons' views with the location, direction and depth of a needle insertion in a biopsy procedure. Source: [Gavaghan et al., 2012].

2.2.2 Visualization system

AR visualization in operating rooms can be provided to surgeons by following devices:

- **projectors:** projecting virtual information onto a patient
- **optical see-through devices:** displaying virtual information on a semi-transparent mirror
- **video see-through devices:** superimposing virtual information on camera views and visualizing through a screen

Projectors. This is a way to display virtual anatomical objects on a patient's body surface to make the reality augmented. A projector can be placed upon a patient or held by a surgeon during surgery (Figure 2.5). The projection technique is based on the registration between a projector and a patient by an external tracking system. Gavaghan *et al.* evaluated the feasibility and usefulness of a mobile projector (Figure 2.5) in 4 clinical scenarios [Gavaghan et al., 2012]:

(1) In open liver surgery, the projector was integrated into a liver navigation system to display a tumor and vessels on a pig liver tissue (Figure 2.6 (A)).

(2) In cranio-maxillofacial surgery, the projector displayed planning data of the mandible, tumor and mandibular branch of facial nerves onto the skin of a patient's left jaw (Figure 2.6 (B)).

(3) For orthopedic oncological surgical navigation, a 3D mesh model of the proximal tibia, a tumor and a safe resection margin was projected onto a cadaver limb via the mobile projector (Figure 2.6 (C)).

(4) In biopsy procedure, the preoperative planning data of a needle's inserting location, direction and depth was projected on a patient's skin via the mobile projector (Figure 2.6 (D)).

Through 4 simulated surgical procedures, the augmented views were found to enable structures to be identified and targeted quickly [Gavaghan et al., 2012]. However, the disadvantages of using projectors are the loss of depth perception and the parallax effect from 2D projection.

Figure 2.5 – Conventional fixed projector, source: [Volonté et al., 2011] and mobile projector, source: [Gavaghan et al., 2012] used for projecting virtual objects onto patients' body.

Optical see-through devices. These devices display both the real-world and the virtual objects on a semi-transparent surface like a half-silvered mirror. The semi-transparent surface reflects the virtual objects meanwhile allows users to see through it. Such as the work presented in [Fichtinger et al., 2005], the reflection of a CT image on a semi-transparent mirror is superimposed on a head phantom which is seen through the semi-transparent mirror, as presented in Figure 2.7(a). Compared to big mirrors, optical see-through head-mounted displays (HMDs) are more often and easier to use, due to that HMD devices generally have light and simple structures and provide natural views. Some examples of such HMD devices are Microsoft HoloLens used in [Liu et al., 2018], NVIS nVisor ST60 used in [Chen et al., 2015], Google Glass and Epson Moverio. Figure 2.7 (b) shows the use of the NVIS nVisor ST60 HMD to visualize the overlay of a virtual bladder model on a pelvis phantom

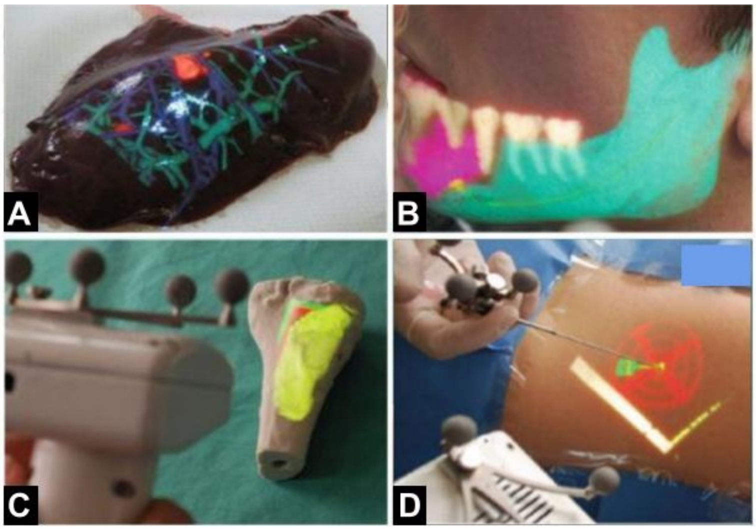


Figure 2.6 – Projecting the virtual information on (A) pig liver tissue, (B) human face, (C) cadaver limb and (D) human skin by a mobile projector. Source: [Gavaghan et al., 2012]

[Chen et al., 2015]. However, optical see-through devices are not used in minimally invasive surgery where operative fields are visualized by endoscopic cameras.

Video see-through devices. In minimally invasive surgery, the operative field is visualized with a video camera (*i.e.* an endoscopic camera) and displayed to surgeons via screens, HMDs or in the consoles of surgical robots, as shown in Figure 2.8. For example, during robot-assisted laparoscopic colorectal surgery, surgeons can see the rectum through endoscopic views provided by a da Vinci® endoscope (Figure 1.5). Therefore, surgeons' views can be augmented by superimposing virtual information on the endoscopic images. The main limitations of these video see-through devices are the restricted field of view (FOV) and the decreased depth perception. A recent study reports that rigid endoscopes typically have FOV between 70° and 110°, with 70° for most laparoscopes and approximately 100° for most arthroscopes [Wang et al., 2017]. It requires a long learning curve for inexperienced surgeons to overcome these limitations. By augmenting the intraoperative endoscopic images with important virtual information, such as virtual tumor model, vessels area and resection margins, AR guidance systems can help surgeons to localize the surgical targets faster.

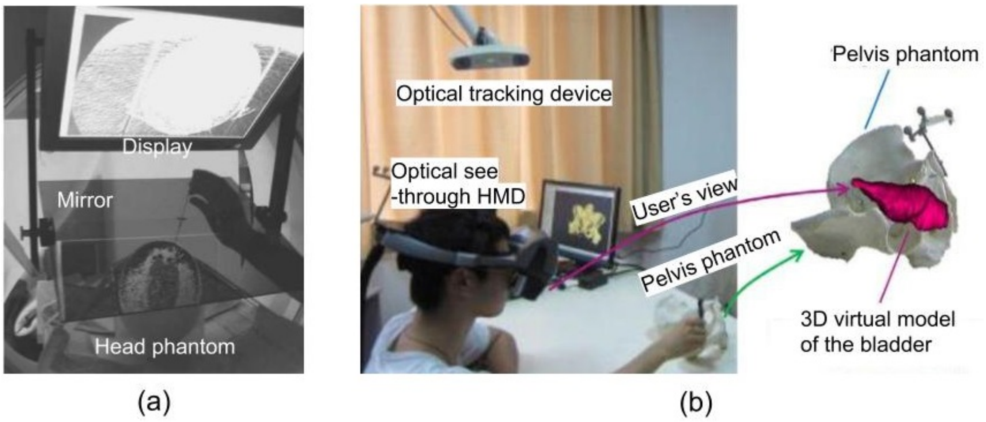


Figure 2.7 – (a) Image overlay system for needle insertion guidance: the display showing the CT image and the reflection of the CT image appearing on the mirror and coinciding with the head phantom behind the mirror. Source: [Fichtinger et al., 2005]. (b) Optical see-through HMD (NVIS nVisor ST60) visualizing the overlay of a virtual bladder model on a pelvis phantom. Source: [Chen et al., 2015].

Conclusion. Projectors and optical see-through devices can be used in open surgery or needle insertion for AR guidance. However, 2D projection makes surgeons miss the depth perception. Moreover, the use of mirrors and lenses in optical see-through devices reduces the brightness and contrast of both real-world and virtual perception. In minimal invasive surgery where an operative field is visualized by a video see-through device, the video augmentation is done before showing to a surgeon, so the brightness and contrast of both real-world and virtual elements can be controlled. Furthermore, video see-through devices such as the da Vinci® visualization system provides high resolution 3D views.

2.2.3 Registration of virtual objects to surgical scenes

Registration is a procedure to align the virtual objects to the endoscopic views of the operative field, in order to provide proper information (*e.g.* accurate tumor location and boundaries) to surgeons. For instance, a 3D mesh model of a rectal tumor should be registered to surgical scenes properly, so that the mesh model would be displayed on the area where the patient's tumor actually is. If a virtual object is created from a patient's medical images (as described in section **2.2.1**), the registration of the virtual object to surgical scenes can be determined by the registration of the patient's medical images to the surgical scenes.

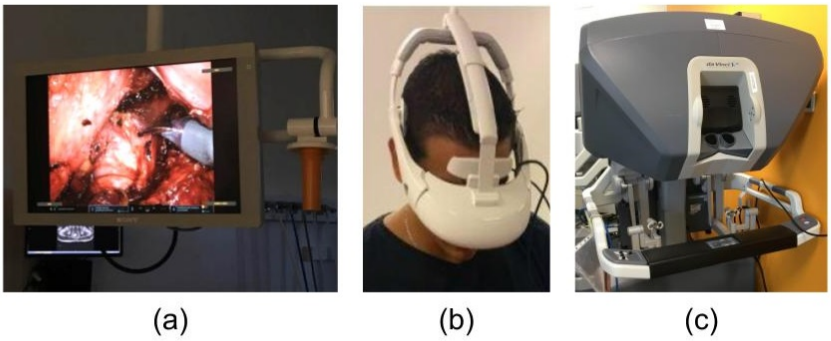


Figure 2.8 – Operative fields can be visualized by (a) a screen, (b) a head-mounted display (HMD) (*e.g.* Sony® HMD) and (c) the console of a surgical robot (*e.g.* da Vinci® surgeon's console) in minimally invasive surgery.

This section discusses state-of-the-art methods to find and maintain a proper registration of medical images to intraoperative endoscopic images (*i.e.* surgical scenes) in minimally invasive surgery. Because of the lack of cross-modality correspondences between medical images and endoscopic images, finding a proper registration between them and maintaining the high level of accuracy of this registration during surgery are the most important and challenging tasks in developing AR systems for intraoperative guidance. The registration can be performed manually or automatically, as introduced below.

Manual registration. The methods of manual registration are simple but slow and highly user-dependent. For example, in order to properly superimpose a virtual tumor model onto intraoperative endoscopic images of operative fields, a surgeon has to manually manipulate the tumor model until a satisfactory degree of anatomical landmarks alignment. Manual registration highly relies on these anatomical landmarks, such as the locations of arteries and veins [Teber et al., 2009]. However, in some surgical procedures, it is difficult for surgeons to identify anatomical landmarks from intraoperative endoscopic views during surgery, such as in transanal endoscopic microsurgery for excision of rectal tumors, as shown in Figure 2.9. Furthermore, in soft tissue surgery, manual registration is lack of accuracy, because neither the tissue deformation under the surface of the operative field nor the organ motion is taken into account.

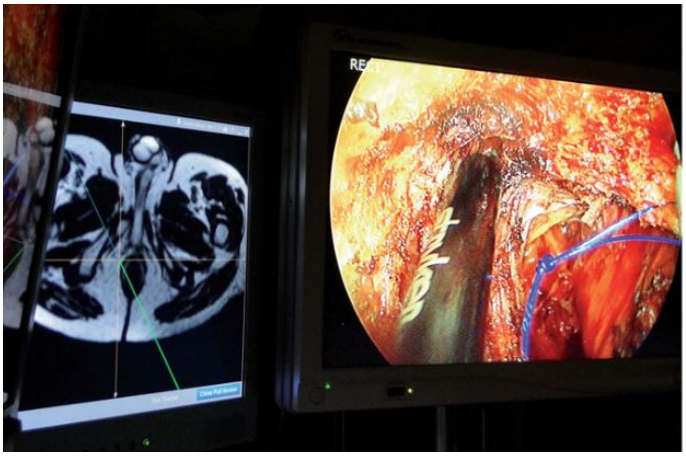


Figure 2.9 – A preoperative MRI image of a patient's pelvis and the intraoperative endoscopic view in transanal endoscopic microsurgery. Source: [Atallah et al., 2016]

Automatic or semi-automatic registration. A number of automatic or semi-automatic registration methods have been proposed and developed based on (1) anatomical landmarks, (2) artificial landmarks, (3) organs' surfaces or (4) intraoperative imaging techniques (*i.e.* volume-based registration) [Bernhardt et al., 2017].

(1) **Anatomical landmark-based registration.** Anatomical landmarks are a set of biologically meaningful point-like features in human bodies. After manually (or automatically) selecting the corresponding anatomical landmarks in both medical images and intraoperative endoscopic views, these landmarks can be automatically matched and their locations in the endoscopic view can be updated via the tracking [Pratt et al., 2012]. However, anatomical landmarks are not always visible in the operative field. Moreover, the registration is based on only the landmarks, so that a large amount of information in operative fields are missed.

(2) **Artificial landmark-based registration.** To alleviate the lack of anatomical landmarks, trackable markers are used as artificial landmarks. During a minimally invasive intervention, some markers are attached to a patient's body surface or an organ's surface before the patient undergoing a medical imaging procedure, and kept in the same location on the patient for intraoperative tracking by the endoscopic camera. These traceable markers are seen as a link between the medical images and the endoscopic images to perform a registration of both imaging modalities. Over the past decade, many kinds of traceable markers have been developed and used for intraoperative tracking: For instance, spherical-shaped chrome steel markers can be tracked by endoscopic cameras

based on their colors [Simpfendörfer et al., 2011]; fluorescent markers (*i.e.* a mix of a Cyanoacrylate, a fluorophore and a contrast agent) are traceable in the presence of smoke or blood in the field of view of the endoscope, and they can remain in patients' bodies after surgery [Wild et al., 2016]; photoacoustic markers are generated on the air-tissue interface by pulsed laser sources, and they are visible in both US images and endoscopic views [Cheng et al., 2014]. However, attaching artificial landmarks to operative fields can be disruptive for the surgical workflow. Furthermore, registration methods based on either artificial or anatomical landmarks have the same problem: they rely on limited information (only landmarks) of the intraoperative scenes, thereby missing the tissue deformation in soft tissue surgery.

(3) **Surface-based registration.** Surfaced-based methods mainly consist of three steps [Marques et al., 2015]:

— **i.** extracting the surface of an organ from medical images by image segmentation and surface rendering;

— **ii.** tracking and reconstructing the surface of the same organ from an endoscopic image;

— **iii.** automatically (or semi-automatically) registering these two surfaces.

However, tracking and reconstructing a surface from an intraoperative endoscopic image remains challenging, because it is difficult to deduce the topography of an intraoperative scene based only on the analysis of some visual cues (*e.g.* edges, conners or other discriminative features) in the endoscopic image. In order to improve the accuracy of surface reconstruction, some studies propose to generate a set of initial points of the surface from an endoscopic image by manual alignment (*i.e.* manually register the surface model to the endoscopic image) [Puerto-Souza et al., 2014] or manually label the region of interest on the endoscopic image [Marques et al., 2015]. Some fully automatic methods to detect organs in laparoscopic images have been proposed, such as the study proposed in [Prokopetc et al., 2015] which used a bounding box to fix the width of an uterus for feature detection. Surface-based methods allow deformable registration based on the information of the surfaces, but the deformation of the inner critical structures beneath these surfaces remains unknown, that is a big problem for intraoperative guidance in soft tissue surgery.

(4) **Volume-based registration.** Intraoperative imaging techniques are able to image the tissues deformation (*e.g.* soft-tissue tumors and vessels) beneath the surface of an operative field during surgery. Some studies use intraoperative images as a common co-

ordinate system, to which preoperative images and intraoperative endoscopic images are registered (*i.e.* volume-based registration). For example, in the study proposed by [Liu et al., 2013], a preoperative CT image of a tongue is registered to its intraoperative CBCT image via elastic registration, then, the CBCT image is registered to the intraoperative endoscopic image, thereby transforming a tumor model from preoperative CT image to the endoscopic image. The registration of the CBCT image to the endoscopic image can be performed by manually aligning the corresponding artificial landmarks [Liu et al., 2013] or by using fluoroscopic images to estimate the endoscope's position in the CBCT coordinate system [Mountney et al., 2014]. In addition to intraoperative CBCT images, intraoperative US images are commonly used for volume-based registration in soft tissue surgery. They will be introduced in detail in section 2.3. A second advantage to use US images is that they show the soft tissues better than CBCT images. This allows to generate the tumor models directly from the intraoperative US images, thereby avoiding to use preoperative images.

Conclusion. Manual registration methods are slow, highly user-dependent and lack of accuracy in soft tissue surgery. The advantages and disadvantages of different automatic or semi-automatic registration methods are shown in the table 2.1. In this thesis, we focus on automatic registration of a 3D intraoperative US image to the intraoperative endoscopic image by using optical tracking technique. It uses as much information of the operative field as possible and avoids complex deformable registration.

2.3 Augmented reality based on intraoperative US

Previously, we stated in section 2.2.1 that intraoperative US is superior to other intraoperative imaging modalities for visualizing soft-tissue tumors and creating tumor models in AR-guided soft tissue surgery. In order to superimpose a tumor model onto an intraoperative scene properly, an accurate registration of the intraoperative US image to the intraoperative endoscopic image is necessary. Figure 2.10 shows a common registration procedure which consists in finding the rigid transformation ${}^{c}\mathbf{T}_{i}$ of the US image (or volume) i to the endoscopic image c. The registration procedure includes 3 parts:

Registration methods	Advantages	Disadvantages
Anatomical landmark-based	biological meaningful features already existing in human body; easy to implement (rigid registration between corresponding landmarks)	using limited information (a set of point-like features) for registration; not always available in operative fields
Artificial landmark-based	a variety of traceable markers available for different surgical procedures; easy to implement (rigid registration between corresponding markers)	artificial landmarks disrupting surgical workflow; using limited information (certain traceable markers) for registration
Surface-based	using more information (organ surface) for registration than landmark-based methods; allowing deformable registration	missing the deformation of the inner critical structures beneath the surface; tracking and reconstructing an organ surface from a small endoscopic view remaining challenging
Volume-based	providing the information of tumor deformation and vessels locations beneath an organ surface for registration	requiring intraoperative imaging systems

Table 2.1 – Advantages and disadvantages of registration methods based on different techniques.

- **US calibration.** It is to determine the rigid transformation ${}^{m2}\mathbf{T}_i$ between an US image i and a marker $m2$ which is attached to the US probe for localization. The US calibration will be introduced in more details in section 3.1 *US probe calibration*.
- **Endoscopic camera localization.** It is to find the rigid transformation ${}^{m1}\mathbf{T}_c$ between the endoscopic images c and a marker $m1$ (which is mounted on the endoscope for localization). In addition to calibrating the endoscopic camera in the coordinate system of the marker, the endoscope can be calibrated in the coordinate system of its robotic holder for tracking and localization [Agustinos et al., 2014]. The camera calibration/localization process will be introduced in more details in section 3.2 *Endoscopic camera localization*.
- **Tracking.** Marker $m1$ and Marker $m2$ are tracked by a tracking system (used as

world coordinate system w) to provide ${}^{w}\mathbf{T}_{m1}$ and ${}^{w}\mathbf{T}_{m2}$. Common tracking techniques includes (1) optical tracking, (2) electromagnetic tracking or (3) computer vision-based tracking [Bernhardt et al., 2017]. In the rest of the section we will introduce these different tracking techniques used in intraoperative US-based AR systems.

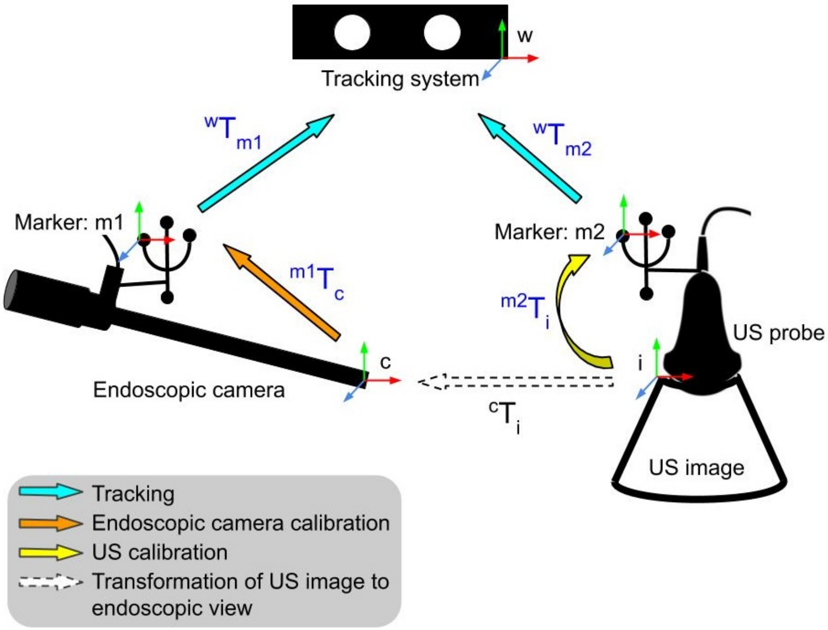


Figure 2.10 – Flowchart of registering a US image to an endoscopic camera.

2.3.1 Optical tracking

This tracking technique uses dedicated cameras to track either active markers (which emit light) or passive markers (which reflect or diffuse ambient light). These markers are attached to the objects that need to be tracked. As shown in Figure 2.10, markers $m1$ and $m2$ are mounted on an endoscope and an US probe, respectively. An optical tracking system is used as the world coordinate system w and tracks the markers $m1$ and $m2$. The registration of the US image i to the endoscopic image c can be estimated by

$$ {}^{c}\widehat{\mathbf{T}}_{i} = ({}^{m1}\mathbf{T}_{c})^{-1}({}^{w}\mathbf{T}_{m1})^{-1}\,{}^{w}\mathbf{T}_{m2}\,{}^{m2}\mathbf{T}_{i} $$

where ${}^{m1}\mathbf{T}_c$ and ${}^{m2}\mathbf{T}_i$ are provided by the endoscopic camera localization and US probe calibration. Figure 2.11 presents a clinical AR application which is implemented based optical tracking technique: In laparoscopic radical prostatectomy, a surgeon uses a transrectal US probe to image a patient's prostate and create a tumor model from the US images, then, the tumor model is projected onto endoscopic views according to the spatial coordinates of the transrectal US probe and the endoscopic camera, which are localized by a Polaris optical tracking system [Ukimura et al., 2010]. The main disadvantage of optical tracking technique is the need for a clear line-of-sight between a tracking system and its markers: markers should be in the field of view of the tracking system without any sort of obstacles between them. This may limit the scope of medical teams' activities in the operating rooms. However, the wireless capability of passive optical tracking technique, the high stability and accuracy have made optical tracking technique generally more favorable in surgical applications than other tracking techniques [Birkfellner et al., 2008; Engelhardt et al., 2016]. Furthermore, recent optical tracking systems offer large working volumes and high level of accuracy, *e.g.* Polaris Spectra® tracking system (Northern Digital Inc.) provides a pyramid working volume with base length 1856 mm, base width 1470 mm and pyramid height 3000 mm, and up to 0.3 mm RMS tracking error within the working volume. For tracking dedicated markers within a pyramid working volume 1327 mm × 1366 mm × 2000 mm, fusionTrack 500® system (Atracsys LLC) provides 0.09 mm RMS tracking error.

2.3.2 Electromagnetic tracking

In some surgical procedures where flexible laparoscopic ultrasonography (LUS) probes are used, electromagnetic (EM) tracking technique is mainly used to overcome the restriction of line-of-sight requirement as optical tracking technique has. Electromagnetic (EM) sensors are small enough to be attached to the tips of LUS probes, thereby localizing the LUS probes inside human body. For example, Aurora® sensor (Northern Digital Inc.) is only 0.3 mm in diameter. As shown in Figure 2.12, two small wired EM sensors are respectively mounted on the flexible tip of a LUS and an endoscope, then, they are positioned above an EM field generator for tracking via an EM tracking box. The calibration of the endoscopic camera and the LUS in the coordinate systems of the EM sensors is required for EM tracking. Figure 2.13 shows the intraoperative use of an EM tracking system. The main advantages of using EM tracking technique in minimally invasive surgery are that it

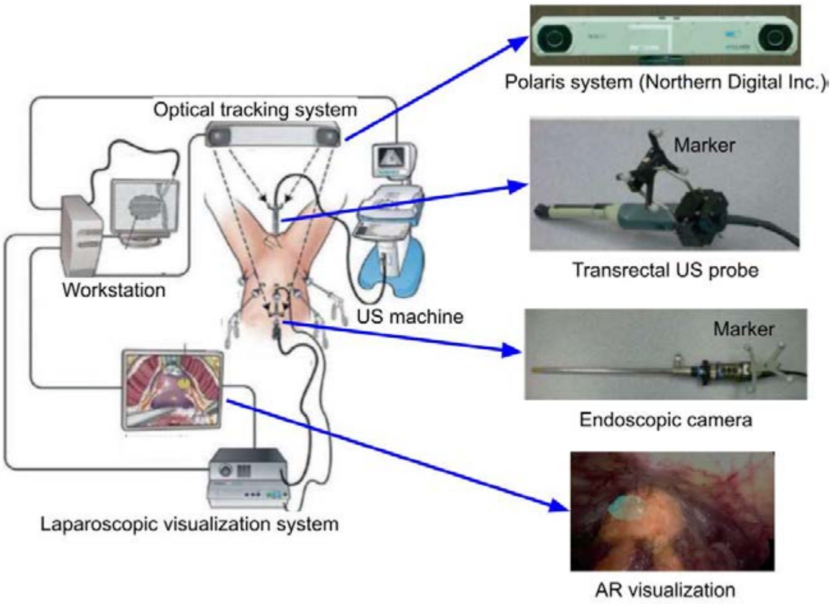


Figure 2.11 – Registration of transrectal US images to endoscopic images using optical tracking technique for projecting the tumor model from US images onto endoscopic images. Source:[Ukimura et al., 2010].

can be used to track flexible instruments, such as LUS probe, endoscope and catheters and it does not have line-of-sight restriction. However, this tracking technique suffers from magnetic field distortions caused by nearby metallic objects, and it provides limited accuracy and smaller working volume compared to optical tracking [Birkfellner et al., 2008; Franz et al., 2014]. For example, one of the most common devices — Aurora® EM tracking system (Northern Digital Inc.) provides up to 1.2 mm RMS tracking error within its working volume 600 mm × 420 mm × 600 mm in an environment without magnetic field distortions.

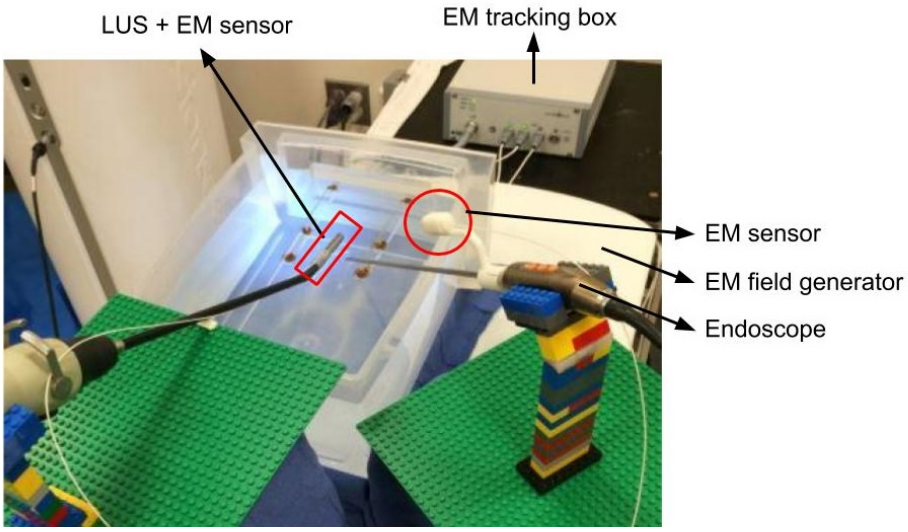


Figure 2.12 – Tracking a flexible laparoscopic ultrasound (LUS) probe and an endoscopic camera via Electromagnetic (EM) tracking technique. Source:[Liu et al., 2016].

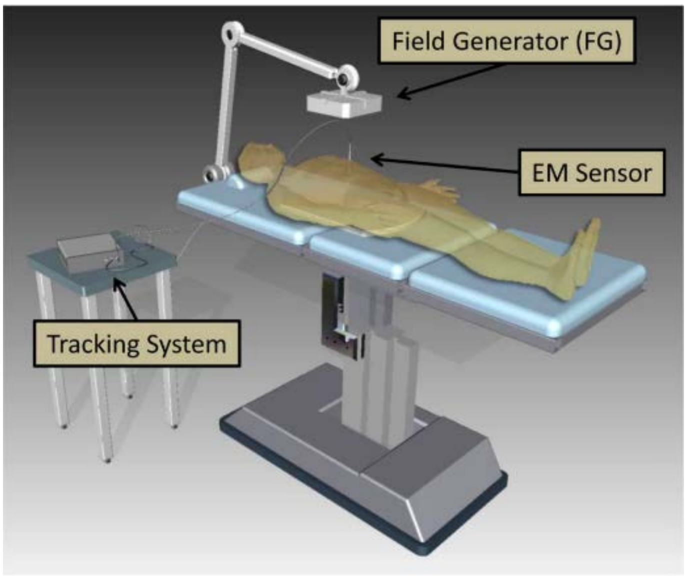


Figure 2.13 – Electromagnetic (EM) tracking technique used in operating rooms. Source:[Franz et al., 2014].

Many studies using EM tracking technique for computer-aided surgery can be found in the literature, as presented in a survey of the state-of-the-art of EM tracking in medicine [Franz et al., 2014]. However, because of different protocols and evaluation methods used in these studies, the results of tracking accuracy are not comparable. Some studies presented relatively high accuracy of LUS-to-endoscope registration using EM tracking systems in their laboratories, where there were no magnetic field distortions in the experimental environments, for example, Cheung *et al.* reported an accuracy of 2.38 ± 0.11 mm [Cheung et al., 2010], and Liu *et al.* reported an accuracy up to 2.59 ± 0.58 mm [Liu et al., 2016]. However, magnetic field distortion is usually unavoidable in an operating room, due to the close proximity of metal objects and electronic equipments [Nakamoto et al., 2008]. EM-optical hybrid tracking techniques have been proposed to predict EM tracking errors caused by the intraoperative magnetic distortion and correct the magnetic distortion in real-time, thereby improving the accuracy of EM tracking in operating rooms [Konishi et al., 2005; Nakamoto et al., 2008; Feuerstein et al., 2009]. Based on the EM-optical hybrid tracking technique, Nakamoto *et al.* presented a vivo experiment using a pig and showed the error (which was caused by intraoperative magnetic distortions) reduced from 44.1 mm to 2.9 mm [Nakamoto et al., 2008]. Moreover, the study of Feuerstein *et al.* showed an EM tracking error reduced from 6.91 mm to 3.15 mm after using an EM-optical tracking system to correct the magnetic distortions [Feuerstein et al., 2009].

2.3.3 Computer-vision-based tracking

Computer-vision-based tracking methods do not need any external tracking systems, but use intraoperative endoscopic cameras to track artificial landmarks in operative fields. These artificial landmarks (*i.e.* markers) are designed to facilitate the detection, so they are more robust and reliable than anatomical landmarks. Commonly, a computer-vision-based tracking procedure consists in following steps:

(1) attaching markers to an operative field, that is shown in intraoperative endoscopic views;
(2) imaging the operative field by intraoperative US;
(3) segmenting the markers from endoscopic images and US images;
(4) registering corresponding markers together;
(5) transforming a virtual model from the US images to the endoscopic images based on the previous registration results.

For example, Simpfendörfer *et al.* proposed to apply this computer-vision-based tracking procedure to a laparoscopic radical prostatectomy by inserting spherical-shaped chrome steel markers into a patient's prostate, as shown in Figure 2.14 [Simpfendörfer et al., 2011]. As a result, the system helped a surgeon to remove the prostate together with the markers. However, in case of big organ deformation and an absent marker, the system is not able to properly augment the endoscopic view based on these markers. Moreover, inserting markers into the operative field can disrupt the surgical workflow.

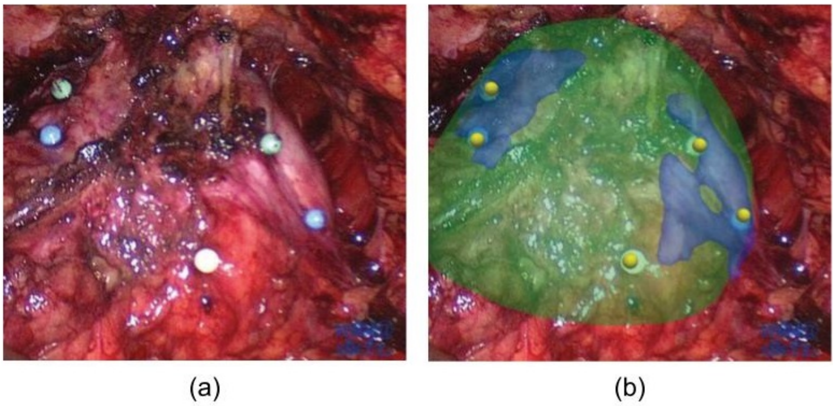


Figure 2.14 – Registration of transrectal US images to endoscopic images based on some spherical markers. Source:[Simpfendörfer et al., 2011].

Another possibility to set landmarks is to use photoacoustic (PA) markers, which are generated on an air-tissue interface without damaging the operative fields. As shown in Figure 2.15, a pulsed laser source generates a PA marker on the surface of a kidney phantom — PA markers are visualized in both endoscopic images and US images, meanwhile, a 3D US probe is used to image the kidney. The US-to-endoscope registration can be performed based on the PA marker [Cheng et al., 2014]. However, the main disadvantage of using a PA marker is the safety concern: in order to generate a PA marker, a laser delivery system has to guide the laser beam into a patient's body and fire it on the surface of an organ. Moreover, setting an appropriate laser energy density seems to be challenging, because different anatomical structures have different capacity to absorb light. For instance, fat has a significantly lower absorption coefficient than blood. Furthermore, 3D US imaging and PA beam-forming should be performed at the same time — which requires technical training, so that the PA marker can be imaged by the US.

Other kinds of markers like camera calibration patterns (*e.g.* chessboard pattern) are also used as markers for tracking and LUS-to-endoscope registration in some recent studies

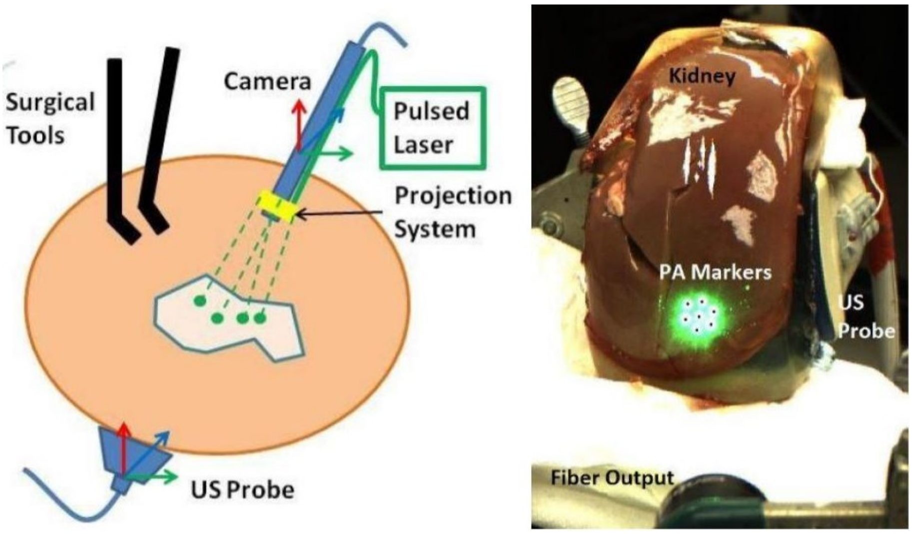


Figure 2.15 – Generating a photoacoustic (PA) marker on an ex-vivo kidney, meanwhile, detecting the PA marker via US imaging. Source:[Cheng et al., 2014].

[Zhang et al., 2017] [Singla et al., 2017]). As shown in Figure 2.16, marker 1 is attached to a laparoscopic ultrasonography (LUS) probe and the US image is calibrated in the coordinate system of marker 1, then, marker 2 is attached to a phantom. The US information can be transformed from the coordinate system of marker 1 to that of marker 2 (*i.e.* the phantom) by the computer-vision-based tracking [Singla et al., 2017]. The main advantage of this method is that these camera calibration patterns are well designed for camera to track patterns' poses, however, the tracking accuracy highly depends on the pattern condition. For example, blood strains in operative fields or rotational motion of LUS probe cause wrong estimation of patterns' poses. Furthermore, tracking accuracy is also related to tracking distance and pattern design (*e.g.* chessboard or circles).

2.3.4 Conclusion

The advantages and disadvantages of the tracking techniques: optical tracking, electromagnetic tracking and computer-vision-based tracking are shown in the table 2.2. This thesis chose optical tracking technique to localize a 3D US probe and the endoscopic camera. Because, firstly, the improvement of optical tracking space helps to minimize the line-of-sight issue. Secondly, it is difficult to avoid the magnetic field distortions from metal objects and

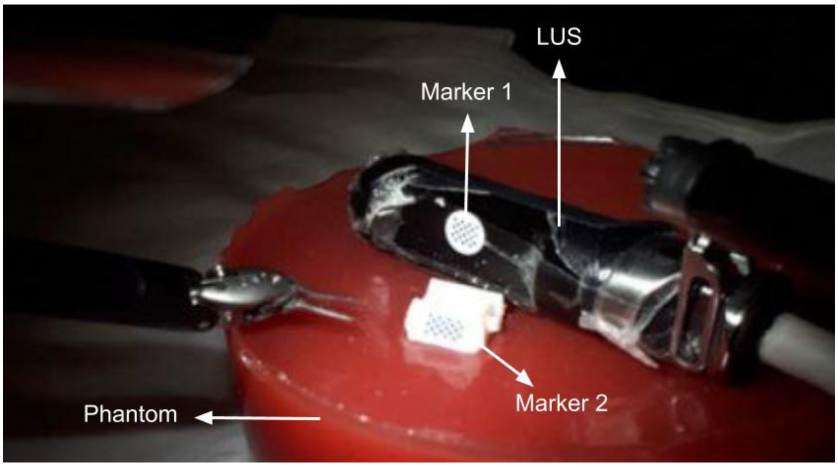


Figure 2.16 – Registering a laparoscopic ultrasonography (LUS) probe to a phantom based on the endoscopic camera tracking some camera calibration patterns. Source:[Singla et al., 2017].

electronic equipments in operating rooms, while using EM tracking technique.

Tracking techniques	Advantages	Disadvantages
Optical tracking	wireless capability; high level of stability and accuracy; large working volumes	line-of-sight issue
Electromagnetic tracking	small EM sensor; no line-of-sight restriction; can be used for tracking laparoscopic US probe	magnetic field distortions; small working volume
Computer-vision-based tracking	no need of external tracking systems	highly relying on artificial landmarks condition; may disturbing surgical workflow; safety concern (PM marker)

Table 2.2 – Advantages and disadvantages of tracking techniques.

2.4 Problem statement and thesis contributions

Robot-assisted minimally invasive surgery allows surgeons to operate with robotic arms, which assist surgeons to perform complex procedures with more precision, flexibility and control than conventional techniques. However, due to the lack of haptic feedback from robotic instruments and small field of view of the endoscopic camera, it remains challenging for surgeons to localize tumors accurately and to define proper resection margins. As introduced in Chapter 1 *Medical context and image-guided systems*, the difficulty of localizing the base of tongue cancer or low-rectal cancer during robotic surgery can be overcome by intraoperative US imaging and AR guidance. Therefore, this thesis proposes to implement an AR framework based on intraoperative US images for intraoperative guidance in robot-assisted surgery — base of tongue cancer and low-rectal cancer surgery. To this end, following problems have to be addressed in this thesis:

- **Visualizing tumor.** Intraoperative imaging technique is required in order to provide tumor information during surgery. Current AR systems mainly use CBCT or/and fluoroscopy for intraoperative localization of base of tongue tumors. However, an ideal intraoperative imaging modality should:
  (1) produce clear images and accurate information of the tumor area;
  (2) be easily installed in the operating rooms for intraoperative imaging;
  (3) be safe to use for both patients and medical teams (no radiation exposure).

  This thesis proposes to use US imaging, which has been commonly used for rectal tumor and base of tongue tumor staging and meets the requirements of being an ideal intraoperative imaging modality. This thesis proposes a fast and accurate US calibration method which significantly facilitates the use of US imaging in the operating rooms Shen et al. [2018].
- **Implementing framework.** We implemented an intraoperative US-based AR framework which augments the endoscopic views with tumor information from intraoperative US images. Finding and maintaining the high level of accuracy of a surgical AR system are the most important and the most challenging tasks, that is determined by the US-to-endoscope registration. Optical tracking technique is chosen to be use in this thesis, due to its reputations of high accuracy, stability, reliability and large working volume. In this thesis, three hand-eye calibration approaches [Tsai and Lenz, 1989] [Daniilidis and Bayro-Corrochano, 1996] [Malti and Barreto, 2010] are implemented and evaluated, in order to accurately track the endoscopic camera. The implemented framework is evaluated with an experimental protocol which is close to clinical scenarios.
- **Simulating integration of AR guidance into surgical workflow.** In order to integrate the proposed framework into the surgical workflows of two specific surgical procedures (*i.e.* transoral robotic surgery for treating base of tongue cancer and

robot-assisted laparoscopic low-rectal cancer), this thesis proposes in collaboration with clinical partners three experimental protocols:

(1) A preliminary experiment with three silicon phantoms to validate the proposed AR framework;

(2) An experimental validation on a ex-vivo lamb tongue by using our proposed AR framework to show the simulated tumor of the ex-vivo lamb tongue;

(3) An experimental validation on a rectum phantom by using our proposed AR framework to guide a colorectal surgeon to identify the resection margins on the rectum phantom.

The main purpose of this thesis is to implement an intraoperative US-based AR framework which can be easily installed in operating rooms.

Chapter 3

Contributions: implementing intraoperative US-based AR framework

Contents

previously, we introduced two specific surgical procedures — transoral robotic surgery for treating base of tongue cancer and robot-assisted laparoscopic surgery for treating low-rectal cancer, based on which we proposed an augmented reality (AR) framework using tumor information in intraoperative ultrasonography (US) images to augment endoscopic views. This chapter presents the procedure of implementing the proposed framework and the used materials. Then, a fast and accurate 3D US probe calibration method, which is developed for calibrating both a swept motor US probe and a transrectal US probe, is introduced in section 3.1. Section 3.2 presents the implementation and evaluation of three state-of-the-art hand-eye calibration methods. Lastly, the implemented framework is evaluated with calibration phantoms, and its performance is shown in section 3.3.

Framework overview. This thesis implements an AR framework based on intraoperative US imaging technique and optical tracking technique. A swept motor 3D US probe and a transrectal US probe, which are currently available in our laboratory, are respectively used for imaging the base of tongue cancer and low-rectal cancer. In this thesis, the swept motor 3D US probe is called 3D US probe, and the transrectal US probe is called TRUS probe. Figure 3.1 shows the framework structure which is separated into two flowcharts in parallel: (1) a visualization flowchart (in purple) that shows what visual information is needed for surgeons and how it is transmitted to the visualization device; (2) a registration flowchart (in orange) that shows how the devices are spatially connected and which spatial transformations have to be estimated.

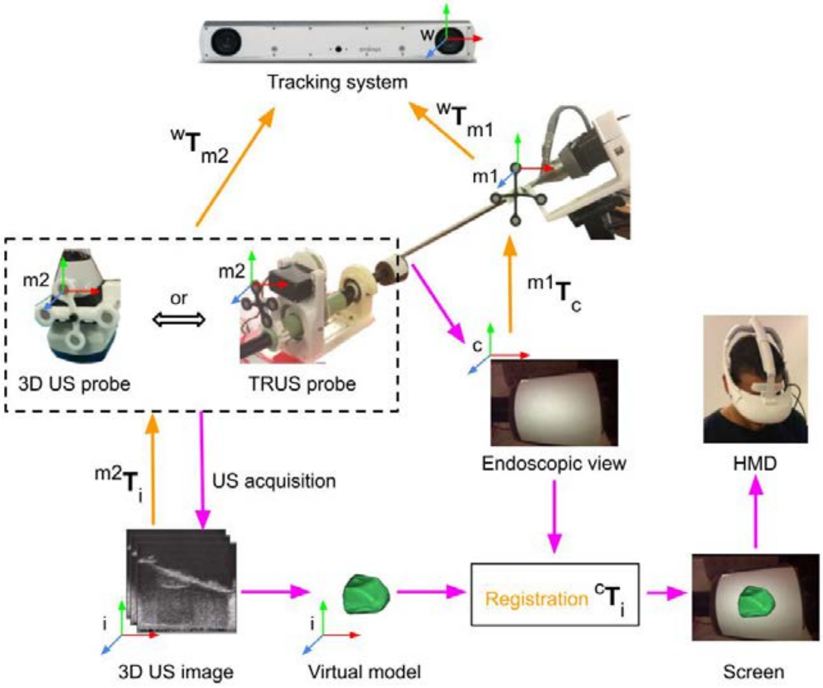


Figure 3.1 – Framework overview: visualization flowchart in purple and registration flowchart in orange.

- **Visualization flowchart (purple).** A 3D US probe or a TRUS probe produces a 3D US image which is segmented to generate a virtual tumor model. Then, the virtual tumor model is registered to, and superimposed on, an endoscopic view. Finally, the augmented view is presented to a user through a head-mounted display (HMD). The HMD device simulates the visualization system of the da Vinci® surgical platform or the visualization system in laparoscopic surgery.

- **Registration flowchart (orange).** The accuracy of an AR application mainly relies on the registration of the 3D US image to the endoscopic view. This registration is performed by using an external optical tracking system with two markers firmly affixed to the endoscope and the US probe, respectively $m1$ and $m2$. In Figure 3.1, i, $m1$, $m2$, c and w represent the coordinate systems of the 3D US image, the marker $m1$, the marker $m2$, the endoscopic camera and the optical tracking system (*i.e.* world coordinate system), respectively. ${}^{b}\mathbf{T}_{a}$ denotes the transformation of coordinate system a to coordinate system b. The tracking system spatially localizes markers $m1$ and $m2$, and provides ${}^{w}\mathbf{T}_{m1}$ and ${}^{w}\mathbf{T}_{m2}$. ${}^{m2}\mathbf{T}_{i}$ is estimated by US probe calibration (section 3.1), and ${}^{m1}\mathbf{T}_{c}$ is computed by the implemented hand-eye calibration methods for the endoscopic camera (section 3.2). According to the registration flowchart (purple), the transformation ${}^{c}\mathbf{T}_{i}$ of the 3D US image i to the endoscopic view c is computed by:

$$ {}^{c}\mathbf{T}_{i} = \left({}^{m1}\mathbf{T}_{c}\right)^{-1} \left({}^{w}\mathbf{T}_{m1}\right)^{-1} {}^{w}\mathbf{T}_{m2} \, {}^{m2}\mathbf{T}_{i} \tag{3.1} $$

Materials. In order to facilitate the clinical implementation of the proposed AR framework, this thesis uses the following materials:

- easyTrack 500 optical tracking system (Atracsys LLC) with Boomerang active markers which are used early in this thesis before having a new version of tracking system;
- fusionTrack 500 optical tracking system (Atracsys LLC) with passive Navex markers which have been used to localize the endoscope and US probes in the world coordinate system;
- SonixTouch Q+ ultrasound system with 4DL14-5/38 Linear 4D ultrasound transducer (a swept motor US probe), which has been used for imaging the base of tongue cancer;
- BK Medical Facon 2101 ultrasound system with 8658 2D transrectal ultrasound transducer (TRUS US probe), which has been used for imaging the low-rectal cancer;
- ENDOCAM Epic 3DHD endoscopy system (Richard Wolf GmbH) which is a similar device used in robot-assisted minimally invasive surgery;
- Sony HMS-3000MT viewing system with head-mounted display certified for medical applications. This device will simulate the consoles of surgical robots which provide visualization to surgeons in operating rooms.

3.1 US probe calibration

3.1.1 Introduction

A 3D US imaging technique is able to provide 3D morphological ultrasound analysis of soft tissues. Two types of 3D US imaging techniques have been proposed to produce 3D US image.

- The first one is called freehand 3D US imaging, which uses a 2D US probe and a position sensor. When a user holds the 2D probe sweeping over a region of interest, a bunch of 2D images are generated to reconstruct a 3D image according to their corresponding positions and orientations (poses) in space. However, the 3D US reconstruction procedure is time-consuming, thereby preventing the freehand 3D US system from performing real-time US volume acquisition. So it is not adapted to our medical context.
- Alternatively, the US volume can be acquired directly by a 3D US probe. Previously, this kind of probes are composed by a transducer mechanically driven by a swept motor. They produce a 3D US image in few seconds (*e.g.* 1.6 seconds by BK Medical® 4DL14-5/38 Linear 4D Transducer). More recently, US volumes can be produced by a 2D matrix array transducer in real-time (*i.e.* volumetric imaging). We will use a 3D probe driven by a swept motor in our applications.

As mentioned in chapter 2, the goal of the 3D US probe calibration is to find the spatial relationship between the 3D US volume and the real word, *i.e.* to localize the position of each voxel of the 3D US volume in the world coordinate system. For this, a marker $m2$ is attached to an US probe for tracking, as shown in Figure 3.1. The produced US image has to be calibrated with respect to the marker $m2$, thereby simplifying the US probe calibration to a procedure of determining the rigid transformation between the 3D US image i and the marker $m2$. Phantom-based US calibration approaches have been commonly used in the literature. As shown in Figure 3.2, a conventional calibration procedure usually consists in:

(1) creating a calibration phantom p, which can be imaged by a US system, and calibrating the phantom with respect to a fixed marker (or a tracking system w) to find ${}^{w}\widehat{\mathbf{T}}_{p}$;

(2) registering the US image i of the phantom to the model of the phantom p to estimate ${}^{p}\widehat{\mathbf{T}}_{i}$;

(3) localizing marker $m2$ of the US probe by a tracking system w to provide ${}^{w}\mathbf{T}_{m2}$

The rigid transformation ${}^{m2}\widehat{\mathbf{T}}_{i}$ of the US image i to the marker $m2$ can then be estimated by:

$$ {}^{m2}\widehat{\mathbf{T}}_{i} = ({}^{w}\mathbf{T}_{m2})^{-1}\, {}^{w}\widehat{\mathbf{T}}_{p}\, {}^{p}\widehat{\mathbf{T}}_{i} \tag{3.2} $$

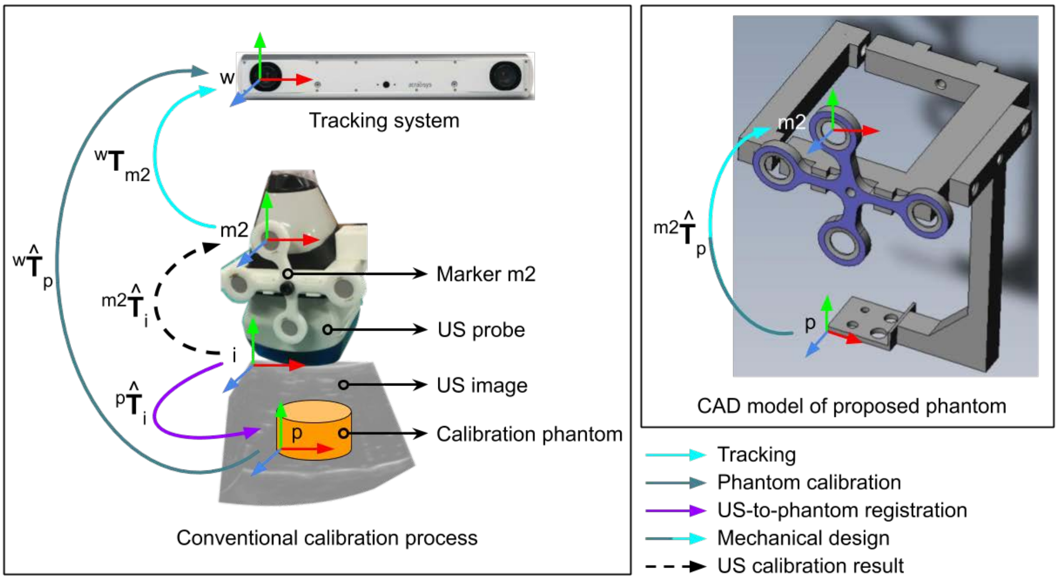


Figure 3.2 – The conventional US probe calibration process to estimate the transformation ${}^{m2}\widehat{\mathbf{T}}_{i}$ (left) and one computer-aided design (CAD) model of a calibration phantom proposed by this thesis.

A comprehensive comparison study proposed in [Bergmeir et al., 2009] showed that phantom-based calibration methods (*i.e.* the phantom is calibrated with respect to a fixed marker or a world coordinate system) provide higher level of accuracy than other methods. A number of studies proposed different calibration phantoms, such as points [Detmer et al., 1994; Hsu et al., 2007], wires [Prager et al., 1998; Bergmeir et al., 2009] or planes [Rousseau et al., 2005; Abeysekera et al., 2014]. In addition to the fact that these phantoms should be build very geometrically precisely, they also suffered from another main disadvantage. They all needed a very precise phantom calibration (*i.e.* calibrating the phantom with respect to a fixed marker), which introduced non-negligible errors into the US calibration process and a relatively long calibration duration even performed by a good metrologist. Abeysekera *et al.* proposed to calibrate multiple US slices before 3D scan conversion process, then, use the best calibration result as the final calibration solution for a swept motor transducer [Abeysekera et al., 2014]. This multiple US slices calibration method shows a high accuracy, but it cannot calibrate 2D matrix array transducers which provide volumetric imaging.

In order to overcome the limitations shown in current studies, this thesis proposes a new calibration phantom which avoids the errors introduced by the phantom calibration. It is a fast and accurate calibration method, really designed to facilitate the intraoperative use of US imaging techniques. Moreover, this automatic calibration method has the ambition to be performed by a non-specialist. Furthermore, the method proposed in this thesis is able to calibrate both swept motor transducers and 2D matrix array transducer. The main ideas of this new calibration procedure is: (1) to use a relatively simple calibration phantom designed in computer-aided design (CAD) and printed using a 3D printer and (2) to fix mechanically this phantom to the US probe and the associated tracking marker m2 in order to avoid the phantom calibration. The CAD model is illustrated in Figure 3.2. The CAD model includes a mounting frame for affixing marker $m2$ to the US probe, and a calibration phantom p for US imaging. Thanks to the mechanical design and the high level accuracy of 3D printing, the phantom p can be precisely fixed on the mounting frame with a known position (*i.e.* the positions of the phantom p and marker $m2$ are unique and constant relative to the mounting frame, so ${}^{m2}\widehat{\mathbf{T}}_p$ is known). Thus, the conventional US calibration procedure (equation 3.2) is simplified to

$$ {}^{m2}\widehat{\mathbf{T}}_i = {}^{m2}\mathbf{T}_p \, {}^{p}\widehat{\mathbf{T}}_i \tag{3.3} $$

Solving a series of problems in conventional calibration procedures is streamlined to solve only the US-to-phantom registration problem, *i.e.* estimating ${}^{p}\mathbf{T}_i$.

3.1.2 Mechanical design of calibration phantoms

In this first study, the Atracsys easyTrack 500 tracking system with a Boomerang active marker was used for tracking, and the 4DL14-5/38 Linear US probe (which is called 3D US probe in this thesis) was used to produce 3D US images. The first design of a calibration phantom was to calibrate the 3D US probe for imaging base of tongue tumors. As shown in Figure 3.3, a Boomerang active marker $m2$ is attached to the US probe by using a mounting frame which is designed and 3D printed to be fixed to the probe firmly. Then, a 3D printed phantom is affixed to the marker $m2$ by embedding the protrusion of $h3$ into $h3'$ and screwing $h1$ and $h2$ to $h1'$ and $h2'$, respectively. After calibrating process, the phantom can be removed from marker $m2$, while, marker $m2$ can be kept at the same position on the US probe for tracking. Based on the CAD models of the Boomerang marker and the phantom, the coordinates of the phantom in the coordinate system $m2$ is known, i.e. ${}^{m2}\mathbf{T}_p$ is known.

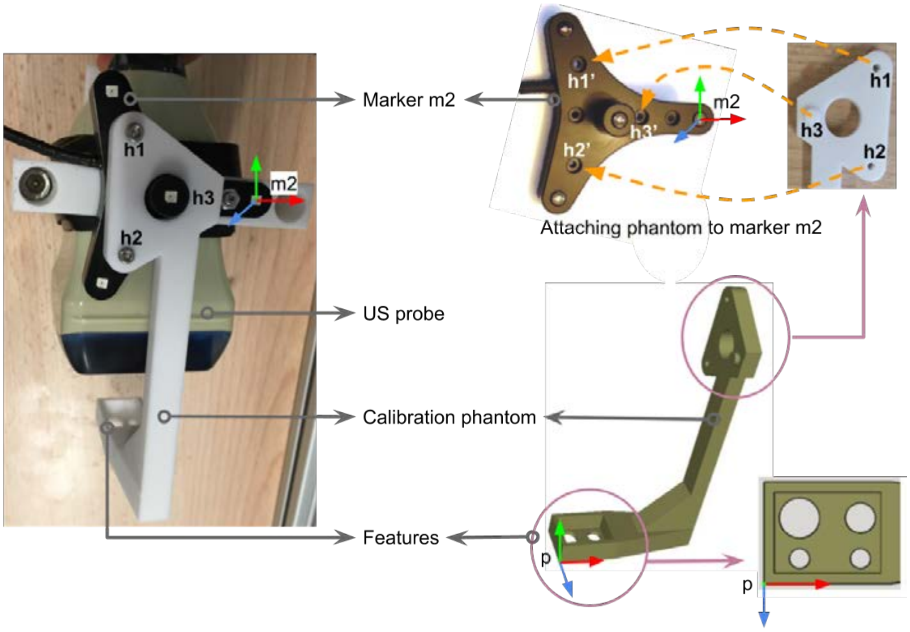


Figure 3.3 – Design of a calibration phantom with respect to an Atracsys Boomerang marker and the 4DL14-5/38 Linear transducer.

In our application, the 3D US probe is used to image the base of tongue tumors through the patient neck. The base of the tongue is generally at a distance of 50 mm to 70 mm from the neck. We designed the shape and the features of our phantom according to this field of view. The features of the phantom, consists of a box shape with a dimension 25 mm × 20 mm × 10 mm. In this box shape, we included holes with radius of 4.0 mm, 3.0 mm, 2.0 mm and 2.0 mm. This box shape and the holes can be easily seen and segmented in the US volume. The dimensions of these features are experimentally chosen based on the field of view of the US probe and the US imaging resolution.

The proposed calibration phantom can be adapted to different markers after few modifications on the CAD model according to the geometries of those markers. For example, the modified calibration phantom in Figure 3.2 is to adapt to an Atracsys Navex passive marker. The phantom's CAD model is 3D printed by the Stratasys Fortus 400mc prototyping machine with Polycarbonates material. Its 3D printing resolution is 0.178 mm.

3.1.3 US image to phantom registration

The transformation ${}^{p}\widehat{\mathbf{T}}_{i}$ (in equation 3.3) is estimated by registering the US image i to CAD model p of the calibration phantom. The registration process consists of 4 steps, as shown in Figure 3.4:

(1) performing US imaging (transmit frequency 10 MHz and depth 75 mm) on the phantom in water to produce a 3D US image of the phantom features;

(2) automatically segmenting the US image by computing image gradient and applying Standard Hough Transform [Gerig and Klein, 1986] on the gradient image;

(3) generating an mesh model from the segmented US image (in orange in Figure 3.4);

(4) registering the mesh model to the CAD model (olive green) of the phantom via surface registration to estimate the transformation ${}^{p}\mathbf{T}_{i}$. The functions of the mesh model generation and surface registration are provided by the 3D Slicer software [Fedorov et al., 2012].

The US imaging parameters were chosen to be compatible with our clinical application for tongue base surgery and can be adjusted for other applications.

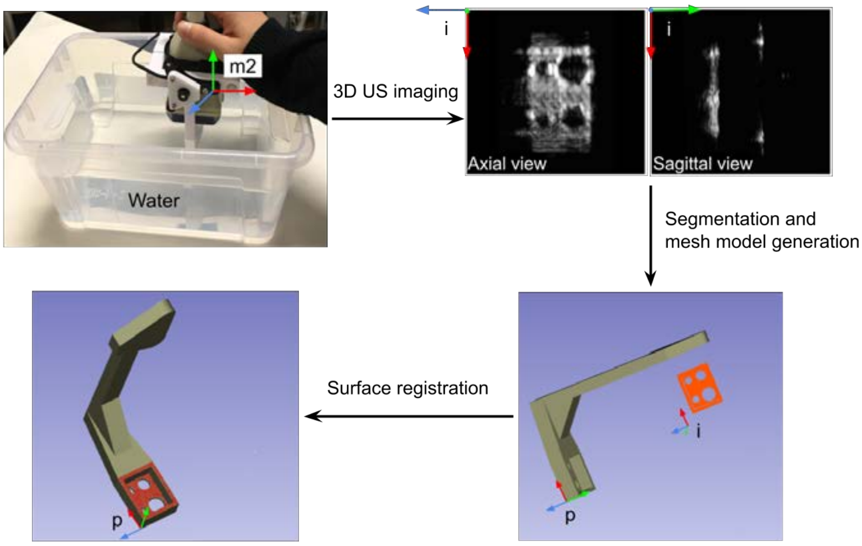


Figure 3.4 – Estimating the transformation between the produced US image i and the phantom p by the registration of two mesh models — the orange model of the US segmentation and the olive green model of the phantom.

3.1.4 US calibration evaluation

3D printing of the calibration phantom may bring errors to the US calibration process. However, the Stratasys Fortus 400mc printer provides high printing resolution and very small printing errors to print a phantom with such size and thickness. The main source of errors in the calibration process comes probably from the registration process between the US image and the calibrated phantom (${}^{p}\widehat{\mathbf{T}}_{i}$). In this section we will evaluate the accuracy of the proposed US to phantom registration method.

Evaluation of US-to-phantom registration. We choose to evaluate the accuracy of the US image to phantom registration (*i.e.* the accuracy of ${}^{p}\widehat{\mathbf{T}}_{i}$ estimation) using the Fiducial Registration Errors (FRE) [Fitzpatrick et al., 1998]. FRE reflect directly the alignment errors induced by the registration, so it is a good marker of the impact of the calibration inaccuracy to the whole framework. In our case, FRE is calculate by:

(1) manually localizing some features (fiducials) in the US image. The proposed phantom is sufficient to extract the features (*i.e.* the corners of the box and the circle centers)

(2) re-projecting these fiducials from the coordinate system of the US image i to that of the phantom p by ${}^{p}\widehat{\mathbf{T}}_{i}$.

(3) calculating the root mean square of distances between the re-projected fiducials and the corresponding fiducials from the CAD model of the phantom.

In order to evaluate the reproducibility of the registration process, the US-to-phantom registration process was performed 10 times by one user, and these estimated $({}^{p}\widehat{\mathbf{T}}_{i})_{n}, n \in \{1,2...10\}$ were evaluated by calculating the FRE. The mean of these FREs was 0.36 mm and the standard deviation was 0.16 mm. Moreover, to verify the effect of the automatic segmentation of the US image on the registration error, we compared it to manual segmentation by: replacing automatic US segmentation with manual segmentation in US-to-phantom registration process; then, performing the registration process 10 times and estimating $({}^{p}\widehat{\mathbf{T}}_{i})_{n}^{manual}, n \in \{1,2...10\}$; finally, calculating the FRE^{manual} using those estimated transformation. The mean of those FRE^{manual} was 0.34 mm and the standard deviation was 0.08 mm. We can see that the automatic US segmentation has similar performance as the manual segmentation (0.36 mm of mean FREs versus 0.34 mm of mean FRE^{manual}).

Point reconstruction tests. Matrix ${}^{m2}\mathbf{T}_i$ represents the rigid transformation between the US image i and marker $m2$ (which is affixed to the US probe). Accuracy of calibrating an ultrasound probe refers to how close the estimated transformation ${}^{m2}\widehat{\mathbf{T}}_i$ is to the real ${}^{m2}\mathbf{T}_i$. To evaluate this accuracy, we run 20 point reconstruction tests using a 3D-printed stylus, as suggested in [Abeysekera et al., 2014]. A point reconstruction test process is shown Figure 3.5 and consists of the following steps:

- A stylus tip is calibrated in the coordinate system w of the tracking system.
- The stylus tip is imaged in water by the 3D US probe and it is segmented manually from the US image (represented by ${}^{i}\mathbf{st}$).
- During the US imaging, a tracking system (with coordinate system w) localizes marker $m2$ of the US probe and the stylus st to provide ${}^{w}\mathbf{T}_{m2}$ and ${}^{w}\mathbf{st}$.
- Point reconstruction error is calculated by

$$e = \mid {}^{w}\mathbf{st} - {}^{w}\mathbf{T}_{m2}\, {}^{m2}\widehat{\mathbf{T}}_i\, {}^{i}\mathbf{st} \mid . \tag{3.4}$$

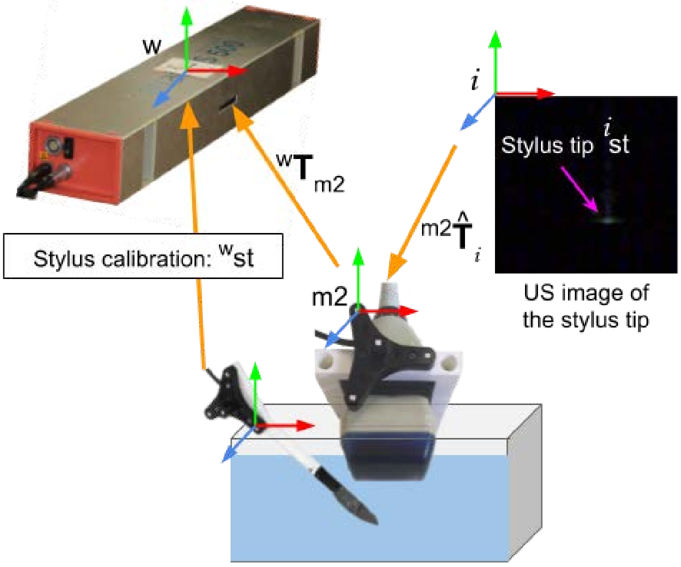


Figure 3.5 – Evaluating 3D US probe calibration by point reconstruction tests using a calibrated 3D-printed stylus.

In addition to small errors from the tracking system and the stylus itself, the smaller point reconstruction error e shows that the ${}^{m2}\widehat{\mathbf{T}}_i$ is closer to the real ${}^{m2}\mathbf{T}_i$. Figure 3.6 shows the boxplot of 20 points reconstruction errors, in which the root mean square (RMS) was

1.39 mm, the mean was 1.26 mm and the standard deviation was 0.62 mm. In these tests, the Atracsys easyTrack 500 tracking system has 0.2 mm RMS error within 1.0 m distance, and the custom-made stylus has 0.83 mm of calibration error.

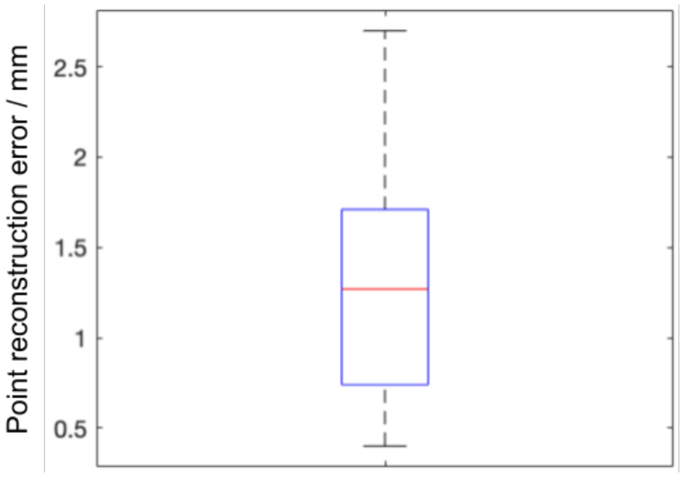


Figure 3.6 – Boxplot of 20 points reconstruction errors.

Comparing calibration performance among different studies is difficult, due to the differences of experimental setup, protocol and environment. Furthermore, each study has its own definition of point reconstruction tests for accuracy evaluation. Some comparison studies have been proposed: For instance, Poon and Rohling compared three phantom-based calibration methods — an IXI-shaped wires phantom, a cube phantom and a stylus, but the evaluation results of those three methods (RMS point reconstruction errors were 2.15 mm, 4.91 mm and 2.36 mm, respectively) are all greater than 2.0 mm [Poon and Rohling, 2005]. Our work achieved a smaller RMS point reconstruction error (1.39 mm) than the errors reported in those comparison studies. However, in point reconstruction tests, the stylus is prone to errors and normally has RMS errors in the range of $[0.6\,\text{mm}, 0.9\,\text{mm}]$ and up to 1.5 mm [Hsu et al., 2007]. The study in [Abeysekera et al., 2014] used a highly accurate stylus with error of 0.11 mm and reported RMS point reconstruction error of 0.93 mm. We suppose that if we had access to such a stylus with lower errors, instead of our custom-made one, our point reconstruction measure would probably be improved.

Conventional calibration methods require not only to track calibration phantoms and US probes, but also to calibrate the phantoms. Those procedures may introduce errors and uncertainties into US calibration process and increase calibration time. This thesis designed a calibration phantom with respect to the marker of the US probe, thereby avoiding the phantom calibration. The fully automatic calibration process can be performed by a non-specialist and developed for intraoperative use of US imaging technique.

In conclusion, as a part of an intraoperative application, a 3D US probe must be calibrated in a fast and simple way with a high level of accuracy. The proposed automatic calibration process is user-friendly and costs less than 5 minutes to be performed by a non-specialist.

3.1.5 Transrectal ultrasound probe calibration

We adapted the calibration method described in the previous section to calibrate the transrectal ultrasonography (TRUS) probe which is used for imaging low-rectal tumors.

3D TRUS imaging system. 3D TRUS are now in common use in urology (robot-assisted biopsy [Vitrani et al., 2016] or brachytherapy of the prostate) and is a good intraoperative device for the localization of the low-rectal tumors, as shown in section 1.2.3. Unfortunately we did not have access to such a 3D probe for our experiments. In order to show a proof-of-concept study with a realistic physical phantom, we decided to create a 3D US imaging system from a 2D TRUS probe which is used in our laboratory. The engineer of our team designed a motorized TRUS probe which automatically collects 2D US images and their corresponding poses (positions and orientations) for 3D US image reconstruction. As shown in Figure 3.7, the probe is driven by a servo motor and rotates along the central axis of the probe. The evaluational resolution (*i.e.* the distance between the neighboring US frames) of the 3D US image is determined by the rectum wall thickness, the TRUS probe diameter and the rotation increment. Overall rectal wall thickness is always smaller than 4 mm (from 2.28 mm ± 0.05 mm to 3.55 mm ± 0.43 mm [Huh et al., 2003]). The TRUS probe used in this thesis has a radius of 10.5 mm. Assuming proper contact between the rectum and the probe, the rectum wall is located inside of the rotation radius $r = 14.5$ mm during US imaging. A rotation increment of $\theta = 2.4^\circ$ is defined to image the rectum wall with a resolution (i.e. distance between neighboring frames) < 1 mm in the field of view of the rectum wall. We fixed a marker $m2$ on the TRUS probe in order to localize the probe in the world coordinate system w and, moreover, to catch more accurate rotation angles θ for the reconstruction of the US volume. When the motorized TRUS sweeps over the region of interest, the collected 2D images and the rotation angle θ are used to reconstruct the US volume with k-Wave MATLAB toolbox [Treeby and Cox, 2010].

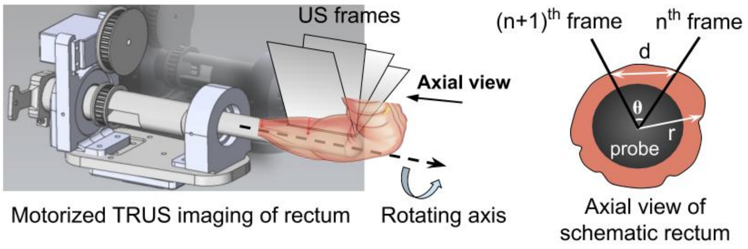


Figure 3.7 – CAD model of the TRUS probe motorization and US imaging of a schematic rectum: d represents the distance between the neighboring frames, θ and r denote the rotation increment and the rotation radius — distance from the probe center to the outer rectum wall.

Phantom design. In the previously described process of calibrating the swept motor 3D US probe (Figure 3.4) used for imaging base of tongue cancer, the calibration problem has been streamlined to a registration of the 3D US image to the CAD model of the calibration phantom. In order to adapt this method to the motorized TRUS probe, the shape of the calibration phantom has been modified as following (Figure 3.8):

- The phantom is tube-shaped in order to surround the TRUS probe, as the rectal wall would do.
- We set some features on the phantom (three circles with the radius of 2 mm, 3 mm and 4 mm respectively, and one 4 mm × 15 mm rectangle) to be imaged by the TRUS system.
- Because the TRUS probe is rotating, we were not able to fix the phantom on the localization marker $m2$ as we had done for the 3D US probe which was used in the tongue experiment. We had to fix a marker $m3$ on the phantom to localize it in the world coordinate system w. Marker $m2$ and marker $m3$ were both tracked by the tracking system.

The tube-shaped calibration phantom was designed on CAD and printed on the 3D printer, as the box-shaped calibration phantom did for the 3D US probe in the tongue experiment.

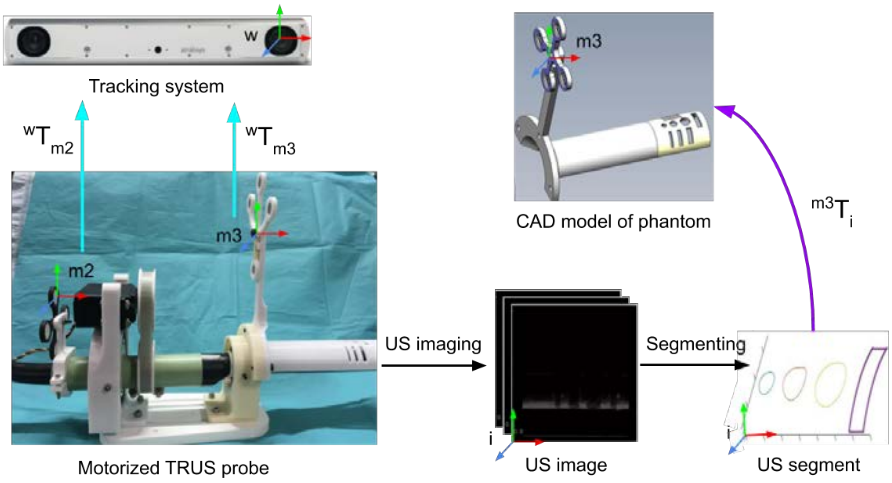


Figure 3.8 – Calibration process using a 3D-printed calibration phantom for transrectal ultrasound (TRUS) probe.

Calibration process. The calibration process is shown in Figure 3.8. It starts with mounting the phantom on the US probe and performing US imaging in water. Next, the features of the phantom in the US volume are automatically segmented by directional gradients and Standard Hough Transform [Gerig and Klein, 1986]. The transformation ${}^{m3}\widehat{\mathbf{T}}_i$ is then estimated by a rigid registration between the US segment and the CAD model of the phantom, as described in section 3.1.3 *US image to phantom registration*. Meanwhile, the tracking system records the data of marker $m2$ and $m3$. Finally, the calibration solution is estimated by

$${}^{m2}\widehat{\mathbf{T}}_i = ({}^{w}\mathbf{T}_{m2})^{-1}\,{}^{w}\mathbf{T}_{m3}\,{}^{m3}\widehat{\mathbf{T}}_i \tag{3.5}$$

After the calibration process, the phantom and marker $m3$ are removed. The accuracy is evaluated with point reconstruction tests (section 3.1.4 *US calibration evaluation*). As shown in Figure 3.9, the point reconstruction errors were computed by equation 3.4 and the mean of 5 points reconstruction errors is 0.9 mm. This error includes the error of 3D US image reconstruction which arises from the motorized TRUS probe sweeping over the region of interest. Considering rectal tumors can grow through the rectum wall and infiltrate the mesorectal fat 15 mm extension, the calibration accuracy was evaluated within the depth of 15 mm.

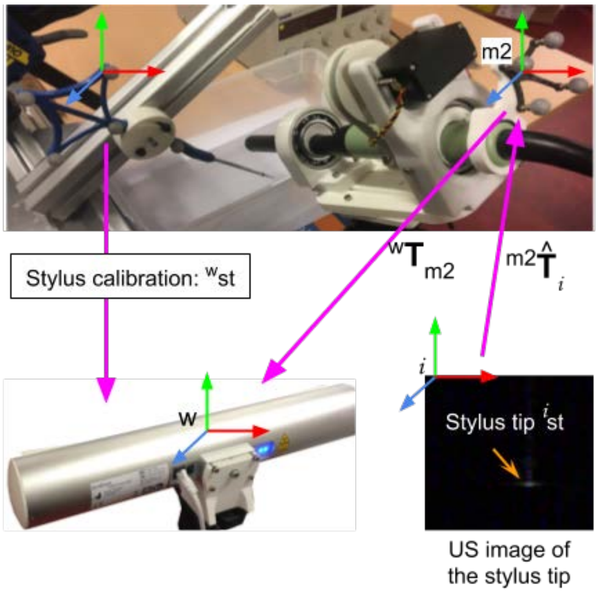


Figure 3.9 – Evaluating TRUS probe calibration by point reconstruction test using a calibrated stylus.

3.2 Endoscopic camera localization

3.2.1 Introduction

Marker $m1$ is affixed on an endoscope for a tracking system to localize the endoscope (Figure 3.1). The endoscopic camera can be calibrated with respect to the marker $m1$ using the hand-eye calibration method, in order to localize the camera in the coordinate system of the tracking system (*i.e.* world coordinate system).

The hand-eye calibration problem first appeared in robotics studies to compute the relative pose between a camera and a robot gripper (the camera is rigidly attached to the robot gripper) [Tsai and Lenz, 1989], as the calibration process shown in Figure 3.10: When the combination of a gripper g and a camera c operate from position $p1$ to position pn, the poses of the gripper are estimated in the coordinate system w of the robot (world coordinate system), so ${}^{w}\mathbf{T}_{g}^{p1}$ and ${}^{w}\mathbf{T}_{g}^{pn}$ are provided. Meanwhile, the camera estimates the poses of a fixed calibration block cb providing ${}^{c}\mathbf{T}_{cb}^{p1}$ and ${}^{c}\mathbf{T}_{cb}^{pn}$. With $\mathbf{A} = {}^{c}\mathbf{T}_{cb}^{pn}\,({}^{c}\mathbf{T}_{cb}^{p1})^{-1}$ and $\mathbf{B} = ({}^{w}\mathbf{T}_{g}^{pn})^{-1}\,{}^{w}\mathbf{T}_{g}^{p1}$, the homogeneous transformation $\mathbf{X}$ (*i.e.* ${}^{g}\mathbf{T}_{c}$) of the camera c to the gripper g can be estimated by solving $\mathbf{AX} = \mathbf{XB}$.

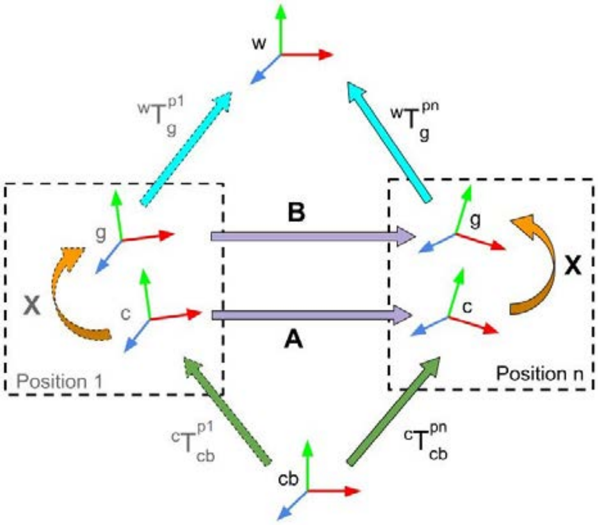


Figure 3.10 – Flowchart of a hand-eye calibration method: g, c, w and cb represent coordinate systems of the robot gripper, the camera, the world (robot work station) and the fixed calibration block, respectively; ${}^{b}\mathbf{T}^{p}_{a}$ represents the rigid transformation from the coordinate system a to the coordinate system b when the gripper and camera operate in position p.

Wengert *et al.* used above calibration process to calibrate an endoscopic camera with respect to a fixed position sensor which is fixed 150 mm from the tip of the endoscope [Wengert et al., 2006]. As a result, Wengert *et al.* presented a calibration error of 3 ± 1 pixels, but did not measure the error in millimeter. In 1999, Daniilidis presented in [Daniilidis, 1999] a hand-eye calibration method which simultaneous calculates the rotation and translation of $\mathbf{X}$ and showed an improved accuracy. In 2010, Malti and Barreto improved Daniilidis's method by using least squares optimization to estimate the translation of the calibration results $\mathbf{X}$, thereby achieving lower translational errors than Daniilidis's method [Malti and Barreto, 2010].

In this thesis, we implement these three methods [Tsai and Lenz, 1989] [Daniilidis, 1999] [Malti and Barreto, 2010], in order to chose the best result to achieve a high level of accuracy in hand-eye calibration for endoscopic cameras.

3.2.2 Camera calibration

Prior to using an endoscopic camera, it is necessary to calibrate it to find the intrinsic, extrinsic and lens distortion parameters. The calibration technique proposed in [Zhang, 2000] has been used with Open Source Computer Vision (OpenCV) Library. In order to adapt this method to the endoscopic camera which has a small field of view, we used a small 13×7 squares chessboard with $1.5\ \text{mm} \times 1.5\ \text{mm}$ for each square. The calibration process is performed within a distance of 60 mm of the endoscopic camera. This distance is realistic relatively to the real operative scene and also present an optimal working space of our endoscopic camera.

3.2.3 Hand-eye calibration

The process of calibrating a 3D endoscopic camera with respect to marker $m1$ is shown in Figure 3.11: Both the left and the right camera of the 3D endoscope capture the chessboard from position 1 to n, and provide ${}^{c}\mathbf{T}_{cb}$ in each position; meanwhile, a tracking system localizes marker $m1$ of the endoscope in each position to provide ${}^{w}\mathbf{T}_{m1}$. These measurements allow to compute

$$\mathbf{A} = ({}^{w}\mathbf{T}_{m1}^{n})^{-1}\,{}^{w}\mathbf{T}_{m1}^{1}$$

and

$$\mathbf{B} = {}^{c}\mathbf{T}_{cb}^{n}\,({}^{c}\mathbf{T}_{cb}^{1})^{-1}$$

and to solve the $\mathbf{AX} = \mathbf{XB}$ problem, where $\mathbf{X}$ represent the hand-eye calibration solution ${}^{m1}\widehat{\mathbf{T}}_{c}$. For a 3D endoscope, the ${}^{m1}\widehat{\mathbf{T}}_{c}$ for the left and right cameras are computed. In practice, it is found that the data acquired in $n = 17$ different positions are sufficient to estimate an accurate ${}^{m1}\widehat{\mathbf{T}}_{c}$.

The three hand-eye calibration methods, proposed in [Tsai and Lenz, 1989] [Daniilidis, 1999] [Malti and Barreto, 2010], were implemented by MATLAB using the dual-quaternion toolbox. These three methods are presented in Appendix. As shown in Figure 3.12, Tsai and Lenz's method (red) provides similar accuracy as that provided by Malti and Barreto's method (blue), while Daniilidis's method (green) seems less accurate. It may be because

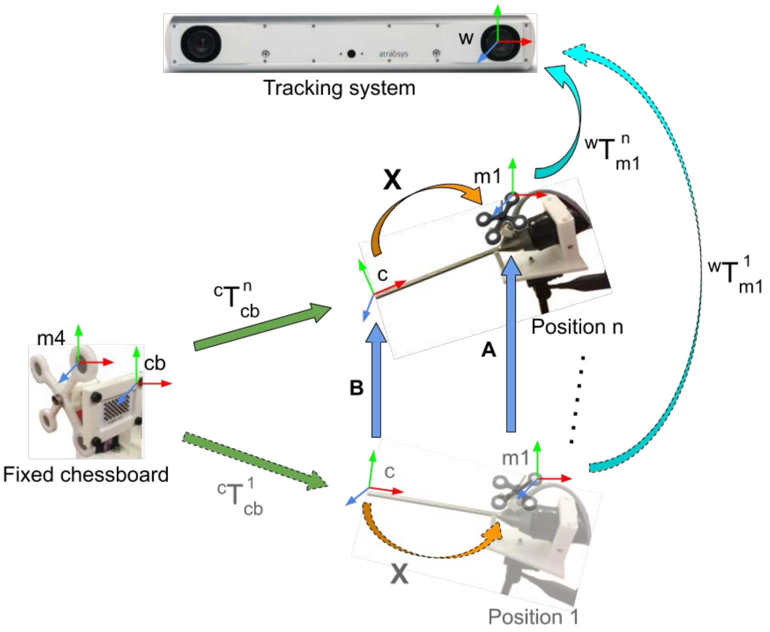


Figure 3.11 – Hand-eye calibration to determine transformation **X** $({}^{m1}\widehat{\boldsymbol{T}}_c)$ between the endoscopic camera and marker $m1$.

Daniilidis's method is not developed for calibrating endoscopic cameras which have small field of views [Malti and Barreto, 2010]. In practice, we chose the results from Tsai and Lenz's method which provides better repeatability and stability during our experiments than Malti and Barreto's method. The calibration is evaluated in section 3.2.4.

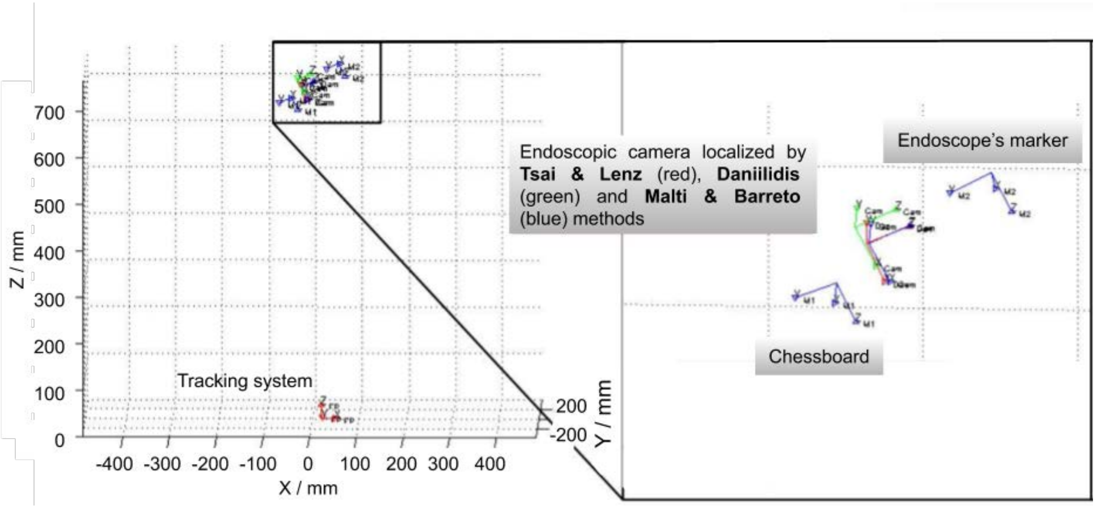


Figure 3.12 – Projection of coordinate systems of the calibration setups.

3.2.4 Evaluating endoscopic camera localization

To evaluate the accuracy of the estimated ${}^{m1}\widehat{\mathbf{T}}_c$ (*i.e.* hand-eye calibration result $\mathbf{X}$), a reflective disk (radius $= 5$ mm) with the coordinate system rd is used and tracked by a tracking system (Figure 3.13). ${}^{rd}P$ is the point cloud extracted from the reflective disk contour in the coordinate system rd. Using the tracking system and the estimated ${}^{m1}\widehat{\mathbf{T}}_c$, the ${}^{rd}P$ is back projected to the coordinate system c of the endoscopic image by:

$$ {}^{c}P = ({}^{m1}\hat{\mathbf{T}}_c)^{-1} \, ({}^{w}\mathbf{T}_{m1})^{-1} \, {}^{w}\mathbf{T}_{rd} \, {}^{rd}P \tag{3.6} $$

${}^{w}\mathbf{T}_{rd}$ and ${}^{w}\mathbf{T}_{m1}$ are the transformation given by the tracking system. ${}^{c}P$ is displayed on the left and right images of the 3D endoscopic camera (green ellipses on the disk in Figure 3.13). Meanwhile, the endoscopic camera captures the reflective disk, as shown in Figure 3.13.

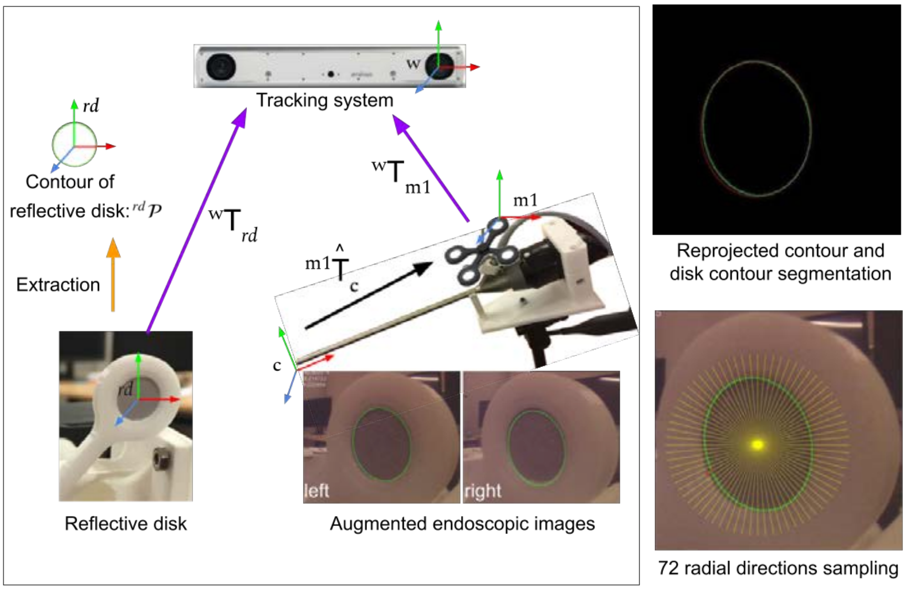


Figure 3.13 – Evaluating the hand-eye calibration solution ${}^{m1}\widehat{\mathbf{T}}_c$: 3D endoscopic camera capturing a reflective disk, and projecting the extracted disk contour (green ellipses) on the left and right camera views.

The accuracy of ${}^{m1}\widehat{\mathbf{T}}_c$ is evaluated by measuring the distance (represented by D_{rms}) between the projected green ellipses and the disk's contour visualized in the augmented endoscopic images. The D_{rms} is measured for the left and right images of the 3D endoscopic camera, respectively. The distance D_{rms} is computed by:

(1) segmenting the green ellipses and the disks contour from the augmented endoscopic images (Figure 3.13) based on the color information of image pixels — the reprojected ellipse is green and the reflective disk captured by the endoscopic camera is grey;

(2) using the previous segmentation to calculate the distances between the green ellipses and the grey contour (disk contour) in 72 radial directions (Figure 3.13);

(3) computing the root mean square (RMS) of these 72 distances, which represents the distance D_{rms} between the green ellipse and the disk contour.

In order to capture the accuracy of the back-projection over the endoscope field of view, we made the endoscopic camera to capture the reflective disk at 12 different poses (position and orientation), which include 6 poses in the field of view of the camera that we called "normal" poses (poses 1 to 6) and 6 poses at the limit of the field of view that we called "extreme" poses (poses 7 to 12), as shown in Figure 3.14 (a). In those extreme poses, the endoscopic camera is either too close (10 mm in pose 7), too far (90 mm in pose 8) or at an extreme side/high/low angle (approximately angle = 45°) at poses 9 to 12 relative to the disk. For each pose, we computed the D^k_{rms} with $k = \{1, 2, ..., 12\}$ represents the pose 1 to 12. The results for D^k_{rms} values are shown in Figure 3.14 (b) and (c). In Figure 3.14 (b), the D_{rms} is smaller than 0.51 mm when the endoscopic camera operates in poses 1 to 6. However, when the endoscopic camera is placed into some extreme poses, the D_{rms} value increases up to 1.52 mm, as shown in Figure 3.14 (c). Therefore, if the endoscopic camera operates in a normal pose, the hand-eye calibration result ${}^{m1}\widehat{\mathbf{T}}_c$ shows a very good performance, which provides errors $D_{rms} < 0.51$ mm.

During the AR framework evaluation and the phantom experiments, the endoscope was placed in an optimal working space, *i.e.* 20 mm to 60 mm away from and $> 45°$ or perpendicular (90°) to the surface of the target object, as in clinical conditions.

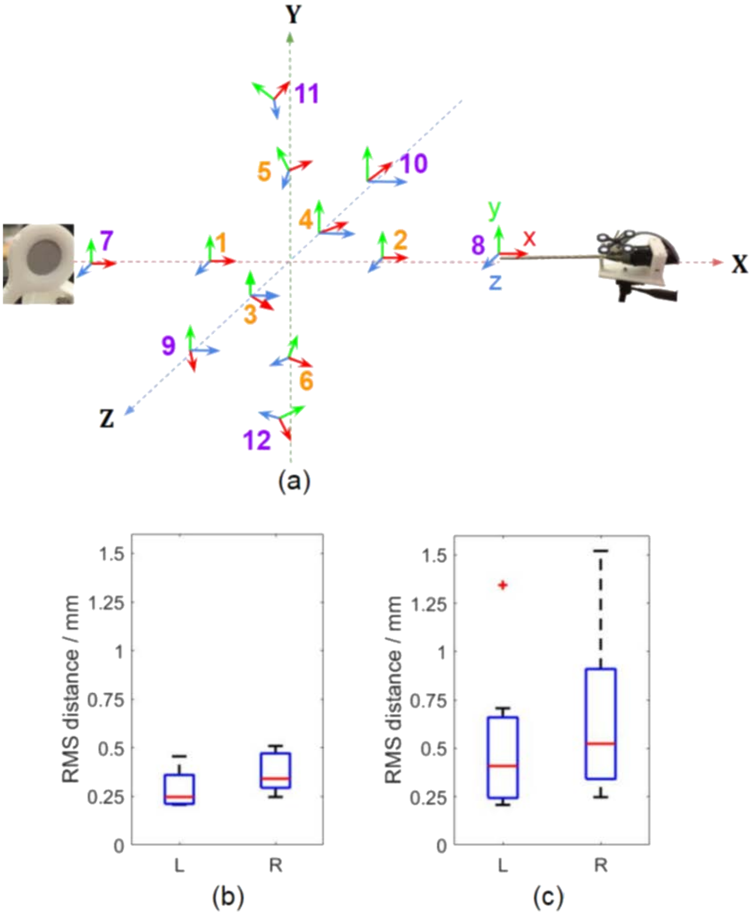


Figure 3.14 – (a) Endoscopic camera capturing the reflective disk in 12 poses (poses 1 to 6 are normal operation poses and poses 7 to 12 are extreme operation poses); boxplot of D_{rms} values for left (L) and right (R) images of the 3D endoscopic camera, obtained when the endoscope operates in normal poses (b) and in both normal and extreme poses (c).

3.3 Framework evaluation

The calibrated ultrasound probe and the calibrated endoscopic camera are tracked by the fusionTrack 500 tracking system (Atracsys LLC) and compose the US-based AR framework. As shown in Figure 3.15, the evaluation process for the 3D US-based AR framework consists in following steps:

(1) attaching the calibration phantom on the 3D US probe (or the TRUS probe in case of

evaluating TRUS-based AR framework);

(2) performing US imaging on the phantom in water;

(3) segmenting the phantom's features from the US image represented by ${}^{i}S$ (the US segmentation of the calibration phantom for TRUS probe is shown in Figure 3.8);

(4) transforming and superimposing the US segmentation ${}^{i}S$ onto the endoscopic camera c according to the equation 3.7, thereby providing ${}^{c}S$ (which is in green in Figure 3.16 and Figure 3.17):

$$ {}^{c}S = ({}^{m1}\widehat{\mathbf{T}}_{c})^{-1}\,({}^{w}\mathbf{T}_{m1})^{-1}\,{}^{w}\mathbf{T}_{m2}\,{}^{m2}\widehat{\mathbf{T}}_{i}\,{}^{i}S \tag{3.7} $$

where ${}^{m2}\widehat{\mathbf{T}}_{i}$ and ${}^{m1}\widehat{\mathbf{T}}_{c}$ are estimated by the US calibration and hand-eye calibration, respectively.

(5) measuring the distance (represented by D_{rms}) between the phantom and the back-projected US features information ${}^{c}S$ on the augmented endoscopic view.

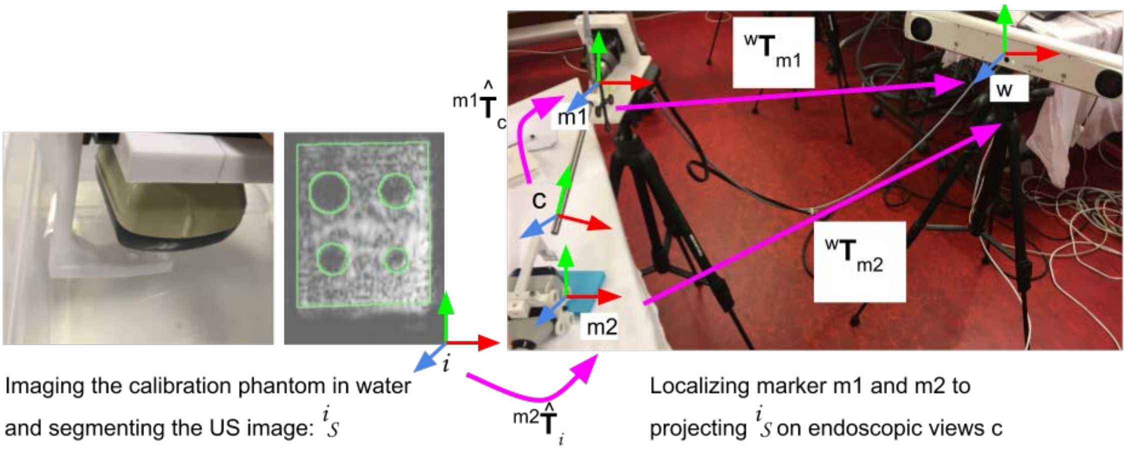


Figure 3.15 – 3D US-based AR framework evaluation: transforming the US segmentation ${}^{i}S$ to the endoscopic view c.

The distance D_{rms} between the green projection and the phantom in Figure 3.16 and Figure 3.17 are computed as described in section 3.2.4 *Evaluating endoscopic camera localization*:

(1) segmenting the projected US features information and the phantom features from the augmented endoscopic images (Figure 3.16 and Figure 3.17) by thresholding;

(2) using the previous segmentation to calculate the distances between the projected US features information and the phantom's features in 72 radial directions;

(3) computing the RMS of the calculated distances, which represents the distance D_{rms} between the green projection and the phantom in Figure 3.16 and Figure 3.17.

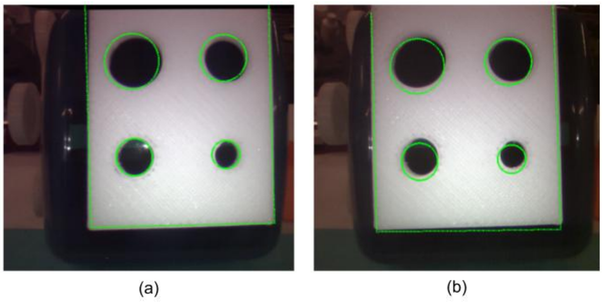


Figure 3.16 – 3D US-based AR framework evaluation: projecting the US segmentation on the left (a) and right (b) views of the 3D endoscopic camera.

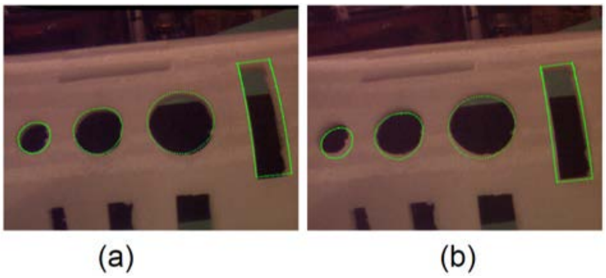


Figure 3.17 – TRUS-based AR framework evaluation: projecting the US segmentation on the left (a) and right (b) views of the 3D endoscopic camera.

In the AR framework using TRUS probe, this evaluation process showed respectively $D_{rms} = 0.45$ mm and $D_{rms} = 0.8$ mm for the left and right endoscopic views. The D_{rms} difference between the left and right camera views is due to stereo camera calibration. The evaluating process was also performed on the AR framework using 3D US probe (4DL14-5/38 Linear 4D transducer), as shown in Figure 3.16. As a result, $D_{rms} = 0.51$ mm and $D_{rms} = 0.87$ mm for the left and right endoscopic views. However, a framework precision less than 1 mm is acceptable in rectal surgery and base of tongue surgery (as confirmed by our clinical partner). The difference of D_{rms} between the left and right endoscopic views is mainly from the stereo camera calibration error.

Chapter 4

Contributions: experimental validation of the framework

Contents

4.1 Introduction

In this chapter, we present three experiments which show the implementation and validation of the proposed augmented reality (AR) framework on physical phantoms. The preliminary experiment was to show the proof of concept of using the proposed AR framework to localize hidden targets in three silicone phantoms. The other two experiments were to simulate the integration of the implemented framework into the surgical workflow of two specific medical applications: transoral robotic surgery for treating tongue base cancer and robot-assisted laparoscopic surgery for treating low-rectal cancer. As a part of the surgical treatment for base of tongue cancer, the AR framework was tested on an ex-vivo sheep tongue. The evaluation of the framework for low-rectal cancer surgery was conducted on a realistic silicone phantom. These two experimental validation were set up in collaboration with our medical partners and allowed them to imagine the contribution of the framework in their future practice.

4.2 Preliminary experiment

This preliminary experiment was to present the proof of concept of using ultrasonography (US) and AR technique to localize hidden targets in three silicone phantoms. In order to estimate the impact of the AR framework compared to the classical framework, we performed two delineation tasks: ***palp*** which mimics the classical procedure using the preoperative image and manual palpation, and ***AR*** which uses our AR framework. Five students from our laboratory and one surgeon participated the experiment.

4.2.1 Experimental setup

Silicone phantoms. In order to simulate tumors growing in soft tissues, we made some silicone objects which mimic soft-tissue tumors, then, put them into silicone liquid to form silicone phantoms, as shown in Figure 4.1. The simulated tumors (*i.e.* silicone objects) can be distinguished from the silicone phantoms by palpation, US imaging or MRI imaging, due to that the simulated tumor is stiffer than the rest of the phantom. As confirmed by our medical partner, the stiffness difference between the simulated tumor and the rest of the silicone phantom is similar as that between a tumor and normal tongue tissues. To simulate the intraoperative deformation of base of tongue tumors during surgery, these silicone phantoms were forced into rigid boxes to deform the simulated tumors (Figure 4.1 (c)). The silicone phantoms in the preliminary experiment were represented by phantoms A, B and C: phantom A had a simulated tumor close to the surface of the phantom, and phantom B and C had simulated tumors in the middle and at the bottom of the phantom, respectively. This was to simulate tumors growing at different depths of base of tongues, thereby providing different tactile feedback in manual palpation task.

(a) Silicone object mimicking a soft-tissue tumor (b) Silicone phantom (c) Phantom placed in a box

Figure 4.1 – Placing a simulated tumor (a) into silicone liquid to form a silicone phantom (b) which mimicking soft tissues, and placing the phantom into a rigid box (c) to deform the phantom.

Task *palp*: localizing simulated tumors via palpation and MRI images.

As described in section 1.1.1 *Surgical treatment for base of tongue cancer*, the classical surgical procedure to identify base of tongue tumors is as follows:

(1) A surgeon prepares the surgical planning using preoperative images (CT or MRI). He/she builds a mental image of the operative field including the information of tumor location and shape, resection margin and surrounding organs.

(2) During surgery, after retracting the tongue for tumor resection, he/she tries to localize the tumor location by manual palpation and uses the information of tactile feedback to reassess the mental image of the operative field and tumor information.

In order to simulate this classical surgical workflow, first, we imaged the silicone phantoms A, B and C by a MRI system, in order to acquire the preoperative images of the simulated tumors before they are deformed. These MRI images are used to help the participants to prepare the delineation of the simulated tumors. Then, we forced these phantoms into the rigid boxes to impose some deformation on the simulated tumors, so that the MR images cannot reflect accurately the locations and shapes of the simulated tumors.

In this task, the goal for the participants was to delineate the tumor shape on the surface of the phantom. We provided the MRI images to the participants and allowed them to palpate these silicone phantoms to assess the locations and shapes of the simulated tumors (Figure

4.2). Then, the participants delineated the boundaries of the simulated tumors by a pointer (i.e. a stylus), as shown in Figure 4.2 (b). The stylus was tracked by a tracking system to provide the trajectory of the stylus tip. In the step of delineating each of these phantoms, the participants used a stereo camera to capture the surface of the phantom and visualize it via a head-mounted display (HMD) to simulate the surgical scenario.

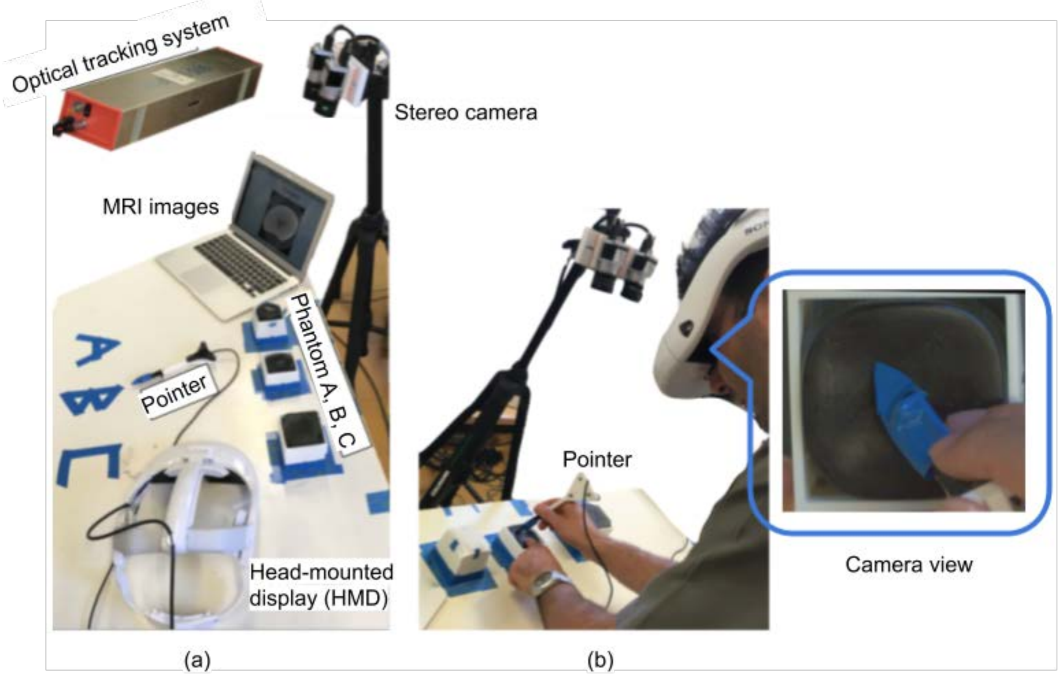


Figure 4.2 – (a) Experimental materials; (b) a participant delineating the boundaries of a simulated tumor in a silicone phantom based on palpation and MRI images.

Task *AR*: localizing simulated tumors via the proposed preliminary AR framework.

Because we did not have access to a 3D endoscopic camera, the preliminary framework was implemented using a pair of industrial cameras which form a stereo camera to provide 3D visualization. Figure 4.3 shows the procedure of calibrating the stereo camera views in the coordinate system of an optical tracking system (*i.e.* world coordinate system w):

(1) The internal and external parameters of the stereo camera can be found by the method of [Zhang, 2000]. So based on the disparity map of the stereo camera c, the camera captures the LEDs of the active marker s providing ${}^{s}\mathbf{T}_{c}$.

(2) Meanwhile, the tracking system w localizes the marker s providing ${}^{w}\mathbf{T}_{s}$.

(3) The transformation ${}^{w}\widehat{\mathbf{T}}_c$ of the camera view c to the world coordinate system (*i.e.* tracking system w) can be estimated by

$$ {}^{w}\widehat{\mathbf{T}}_c = {}^{w}\mathbf{T}_s \, {}^{s}\mathbf{T}_c \tag{4.1} $$

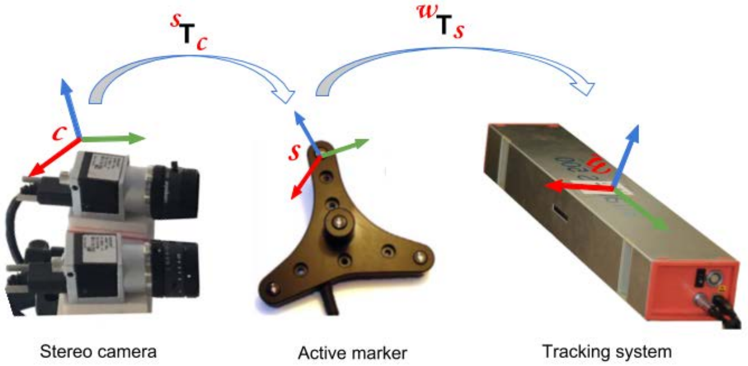


Figure 4.3 – Calibrating a stereo camera c in the coordinate system w of the tracking system using an active marker s

.

After the calibration of the US probe (estimating ${}^{m2}\widehat{\mathbf{T}}_i$ in section 3.1) and the stereo camera with respect to the world coordinate system w (*i.e.* the tracking system), the information in the US image i can be transformed to the coordinate system c of the camera views by applying the following transformation on it:

$$ {}^{c}\widehat{\mathbf{T}}_i = ({}^{w}\widehat{\mathbf{T}}_c)^{-1} \; {}^{w}\mathbf{T}_{m2} \; {}^{m2}\widehat{\mathbf{T}}_i \tag{4.2} $$

Figure 4.4 illustrates the procedure of augmenting the stereo camera views with the virtual models of the simulated tumors. It consists of following steps:

(1) segmenting the simulated tumor from MRI images by thresholding for creating a virtual model by 3D Slicer software [Fedorov et al., 2012];

(2) performing US imaging on the phantom;

(3) performing MRI to US registration to estimate the transformation ${}^{i}\widehat{\mathbf{T}}_{mri}$;

(4) transforming the virtual model from the coordinate system mri of the MRI image to that of camera views c by

$$ {}^{c}\widehat{\mathbf{T}}_{mri} = {}^{c}\widehat{\mathbf{T}}_i \; {}^{i}\widehat{\mathbf{T}}_{mri} \tag{4.3} $$

where ${}^{c}\widehat{\mathbf{T}}_i$ is estimated in equation 4.2.

(5) superimposing the virtual model on the camera views (Figure 4.5).

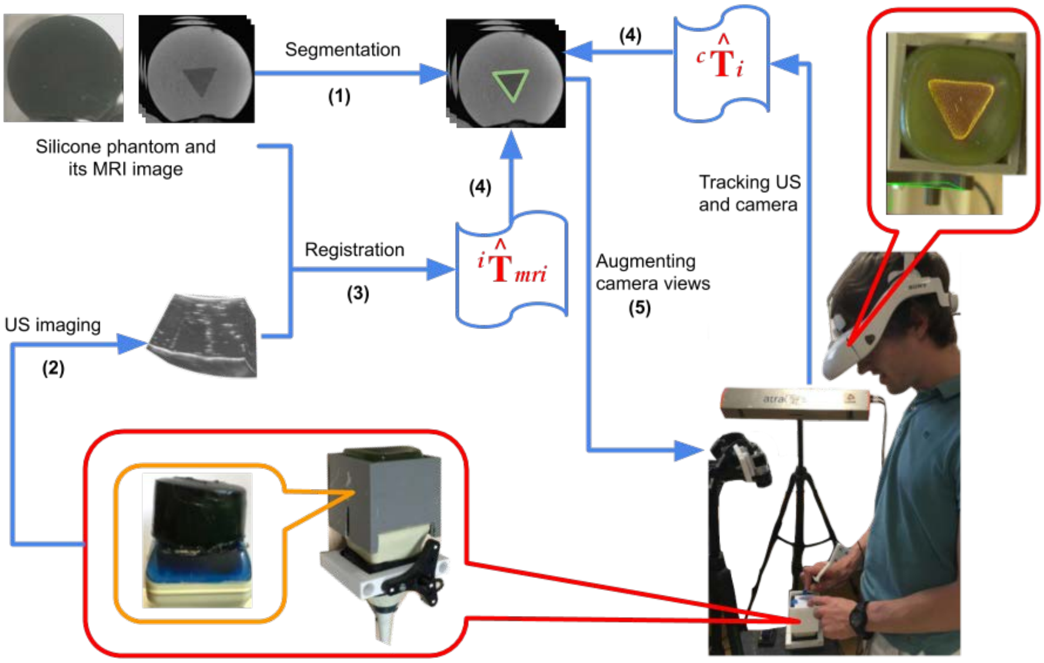


Figure 4.4 – Procedure of implementing a preliminary US imaging-based AR framework.

The goal assigned to the participants in the ***AR*** task is similar as in the ***palp*** task, but the participants had to use our AR framework. According to the augmented views (Figure 4.5), each of the participants (*i.e.* five students and one surgeon) delineates the simulated tumors on the surfaces of the phantoms A, B and C using a stylus. The stylus was tracked by the tracking system to provide the trajectory of the stylus tip.

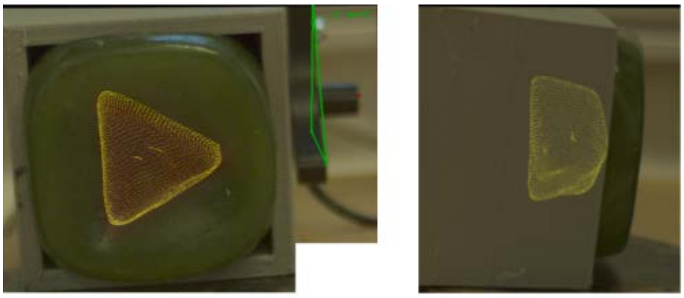

(a) Front view
(b) Side view

Figure 4.5 – Augmented camera views.

4.2.2 Results of the preliminary experiment

Phantoms A, B and C were used in the experiment. Each of these phantoms was placed firmly on the US probe by using a rigid box (Figure 4.6). Then, the US probe was fixed on a desk during the tasks of ***palp*** and ***AR***. Just before a participant delineating a phantom, the ground truth of the phantom has to be generated for comparison of the participant's delineation to it, as shown in Figure 4.6. It consists of following steps:

(1) using a calibrated stylus to localize the corners of the box to create a plane ${}^{w}Pc$ in the coordinate system w of the tracking system;

(2) performing US imaging on the phantom and segmenting the simulated tumor from the US image by thresholding;

(3) transforming the US segmentation to the coordinate system w of the tracking system by applying ${}^{w}\mathbf{T}_{m2}\ {}^{m2}\widehat{\mathbf{T}}_{i}$ (previously estimated by US calibration in section 3.1) on it: ${}^{w}Seg$;

(4) projecting the US segmentation ${}^{w}Seg$ to the plane ${}^{w}Pc$ to generate the ground truth for the phantom (Figure 4.6).

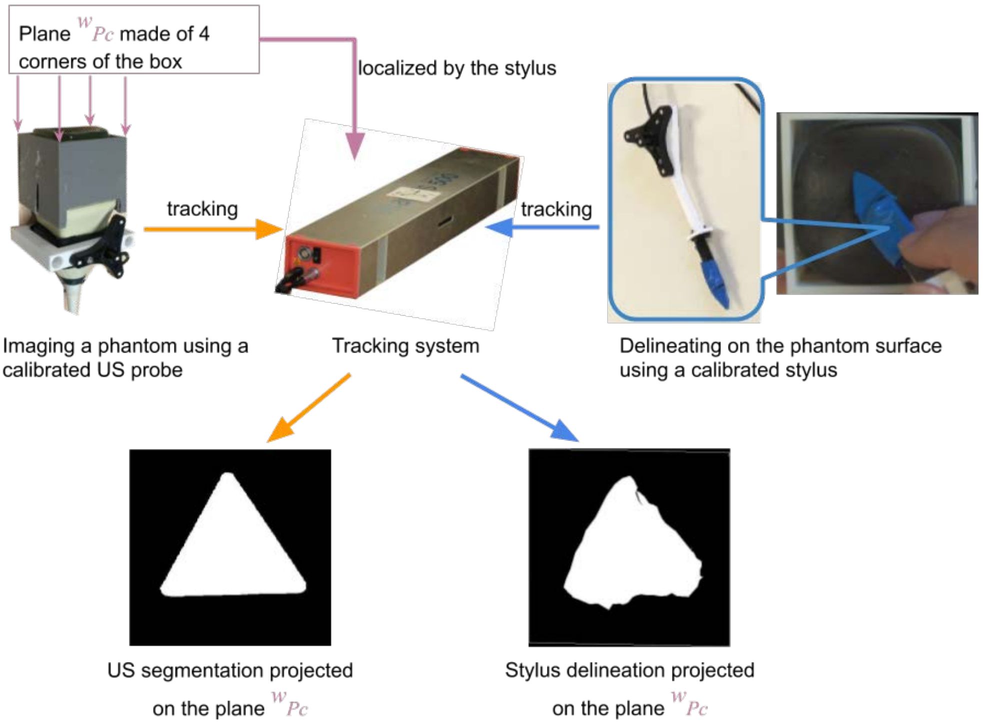


Figure 4.6 – Generating ground truth and participant's delineation for comparison.

Then, a participant delineated the simulated tumors on the surfaces of these phantoms using a stylus. The stylus tip was calibrated with respect to the coordinate system w of the tracking system, so that the trajectory of the stylus tip can be recorded by the tracking system. Figure 4.6 shows an area of the stylus delineation projected onto the plane ${}^{w}Pc$ of the phantom box, so that this projection can be superimposed onto the ground truth of the corresponding phantom for comparison. Figure 4.7 presents superimposing a participant's delineations (green) onto the ground truth (purple) of the phantom, where the overlapped area is represented by white color.

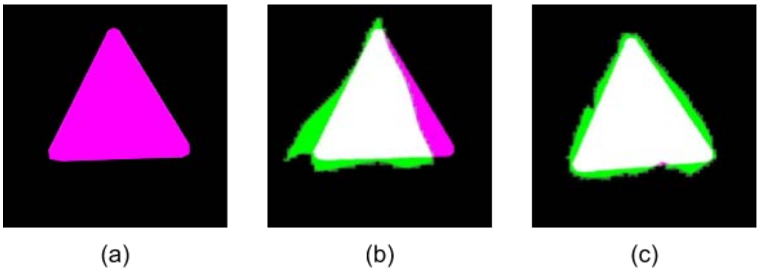


Figure 4.7 – Superimposing a participant's delineations from the task ***palp*** (b) and the task ***AR*** (c) onto the ground truth of the phantom (a).

Two criterions (Dice coefficient [Dice, 1945] and a modified Hausdorff distance [Dubuisson and Jain, 1994]) were used to evaluate the overlapping of the participants' delineations and the ground truth of corresponding phantoms (such as Figure 4.7 (b) and (c)). Figure 4.8 (a) illustrates the results of the Dice coefficient, in which we can see that the Dice coefficient remains clearly higher in ***AR*** than in ***palp***. This indicates that the participants' delineations from the task ***AR*** is closer to the ground truth than these from the task ***palp***, as it is expected. Moreover, the most professional participant (the surgeon) always obtained the higher Dice coefficient in the task ***palp*** than other 5 participants (the students from our laboratory). However, two students get higher Dice coefficients than the surgeon in the task ***AR*** when they delineated the phantom A and B. The Hausdorff distance is shown in Figure 4.8 (b), in which the distance obtained from the task ***AR*** is much smaller than that from the task ***palp***. The surgeon obtained the smallest Hausdorff distance 3.27 mm in the task ***palp***, but one student got 0.71 mm in the task ***AR***.

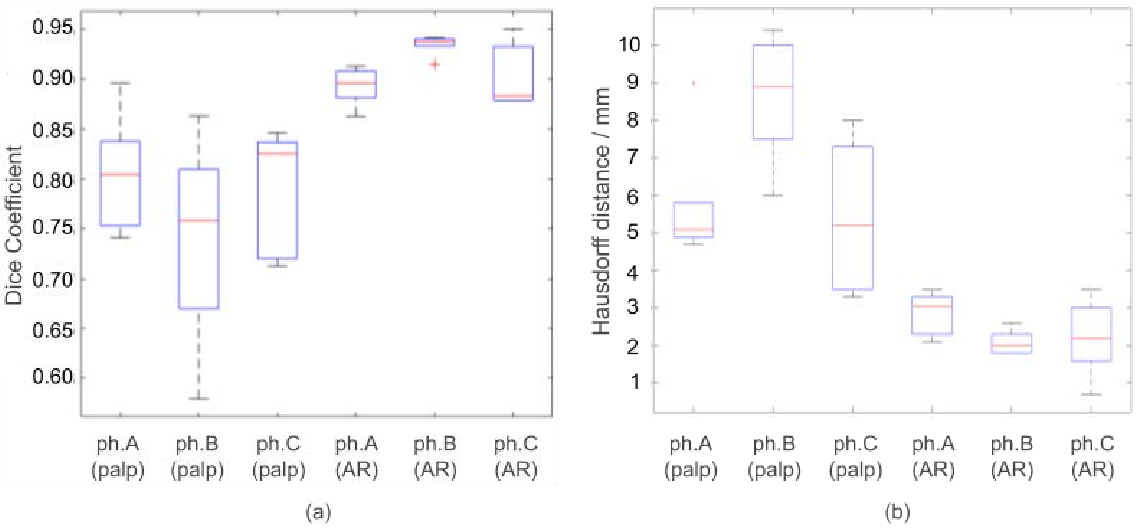


Figure 4.8 – (a) Boxplot of the Dice coefficients (1 represents 100% overlap, and 0 means 0% overlap) and (b) Hausdorff distance: ph.A, ph.B and Ph.C represents silicone phantoms *A*, *B* and *C* in the task ***palp*** and the task ***AR***, respectively

4.2.3 Discussion of the preliminary experiment

This preliminary experiment is to show the proof of concept of implementing an AR framework based on US imaging technique for localizing hidden simulated tumors in silicone phantoms. As it was expected, the participants easily localized the simulated tumors and provided more accurate locations of the simulated tumors according to the AR guidance than that according to palpation and MRI images.

In the experiment, as a part of the preliminary AR framework, a pair of industrial cameras was used to form a stereo camera. It is replaced by a 3D endoscopic camera in the following two experiments, in order to make the experiments as close to the real surgical procedure as possible. After the preliminary experiment, we discussed with surgeons and realized that the intraoperative US technique, as a sole imaging technique, is able to show soft-tissue tumors in US images. So we decided not to use MRI images in the following validation experiments. However, it is possible to integrate tumor information in MRI images to the proposed framework by deformable MRI to US registration.

4.3 Validating the framework on an ex-vivo lamb tongue

This experiment is to evaluate the performance of the developed framework for the guidance of a tumor resection in the context of base of tongue surgery. We mimic this gesture on an ex-vivo lamb tongue. Moreover, our goal was to prove the feasibility of integrating the US-based AR guidance into the surgical workflow of the base of tongue cancer surgery.

4.3.1 Experimental setup

In order to be as close as possible to the actual clinical condition, we simulated a tumor in an ex-vivo lamb tongue. Figure 4.9 (b) presents the ex-vivo lamb tongue (with thickest part approximately 35 mm). The simulated tumor was made of relatively stiff silicone and its size was approximately 10 mm × 8 mm × 5 mm. Then, we made a small open wound on the left side of the tongue, inserted the simulated tumor into the tongue, and sutured the wound.

We implemented the AR framework for the tongue experiment as follows (Figure 4.9 (a)):

(1) The 3D US probe and the endoscopic camera were calibrated as described in the previous chapter 3. The fusionTrack 500 optical tracking system (Atracsys LLC) localizes the US probe and the endoscope in the operating space.

(2) In the actual clinical condition, the 3D US imaging is performed on the patients' submental areas. However, it was difficult to find an ex-vivo tongue with a submental part. We placed the 3D US probe under the tongue. To avoid the movements, we fixed the probe on a table and placed the tongue in a stable manner over the probe. The tumor model was clearly perceptible in the US image (Figure 4.9 (a)).

(3) The participant manually segmented the tumor model in the US volume and created the virtual tumor model by 3D Slicer software [Fedorov et al., 2012]. The fixed setting of the US probe gave us the time to perform the segmentation. In the future, an automatic segmentation will help to segment the tumor in real time in order to recover the dynamics of the probe and organ.

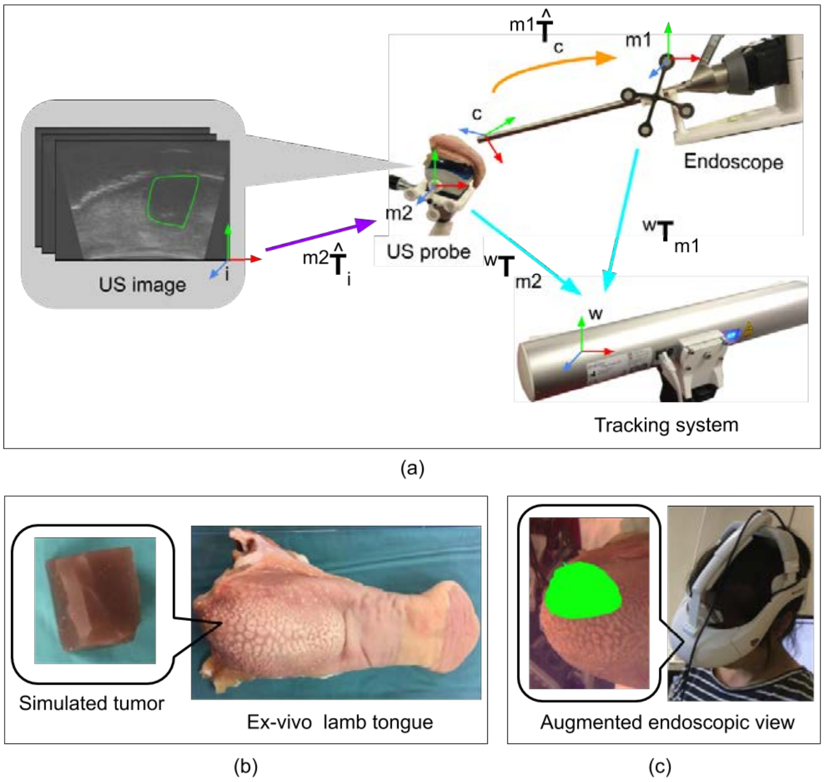


Figure 4.9 – (a) Registration process to augment the endoscopic views with the US segment of the simulated tumor; (b) inserting a simulated tumor in an ex-vivo lamb tongue; (c) displaying augmented endoscopic views on head-mounted display (HMD)

(4) Based on the registration results between the US image and the endoscopic camera, as explained in chapter 3, the virtual tumor model was superimposed onto the endoscopic views of the tongue. It was displayed to the participant by the HMD (Figure 4.9 (c)). By moving the endoscopic camera around the tongue, we can find a proper view for the participant, that is approximately 40 mm between the endoscopic camera and the tongue surface. The endoscopic camera was placed around 70° to the surface of the tongue, where the endoscopic camera perpendicular to the tongue surface is represented by 90°.

(5) The participant wearing the HMD marked the resection margins directly on the tongue with a marker pen. The resection margin was marked at approximately 10 mm away from the boundaries of the virtual tumor model. This step was to mimic the step of a surgeon marking the resection margins by slightly cauterizing the tissues with electro-surgical devices during surgery.

(6) Finally, the participant resected an area of the tongue following the marked margins. Indeed, he cut the marked area with a scalpel perpendicular to the tongue surface.

In this experiment we did not take the depth of the tumor into account. The main objective of the experiment was to evaluate if the AR framework was able to show precisely the projected shape on the tongue surface.

4.3.2 Results of the ex-vivo lamb tongue experiment

To evaluate the resection, the resected specimen was imaged in water via US imaging, as shown in Figure 4.10. We used 3D Slicer software [Fedorov et al., 2012] to visualized the US volume of the resected specimen in 3D. The axial images of the US volume show both the boundaries of the simulated tumor and the resection margins in the axial plane of the specimen (*i.e.* the axial plane of the tongue). The depth of the resection was not take into account, because the goal of this experiment was to evaluate the performance of the AR framework projecting information on the tongue surface. On each of 10 axial US frames (with 1 mm between neighboring axial US frames), we measured the the maximal and minimal distances between the 'tumor' boundaries and the specimen edge, as shown in Figure 4.11. Figure 4.12 illustrates the maximal and minimal distances measured on 10 axial US images. It shows that the mean maximal distance is 11.7 mm (1.7 mm bigger than the defined resection margin) and the mean minimal distance is 8.5 mm (1.5 mm smaller than the defined resection margin).

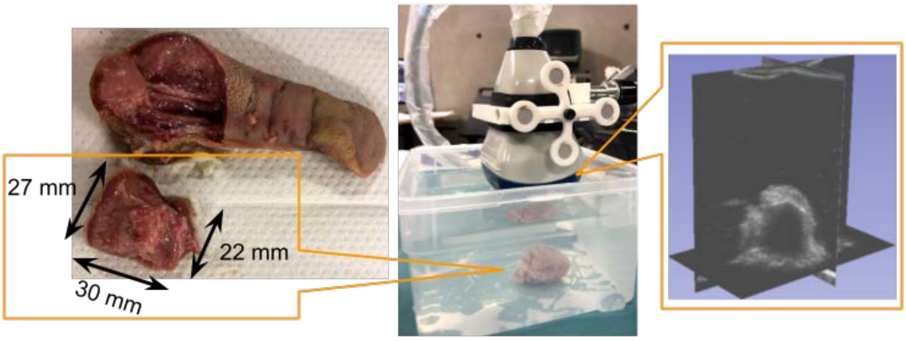


Figure 4.10 – Evaluating the resection of simulated tumor by US imaging.

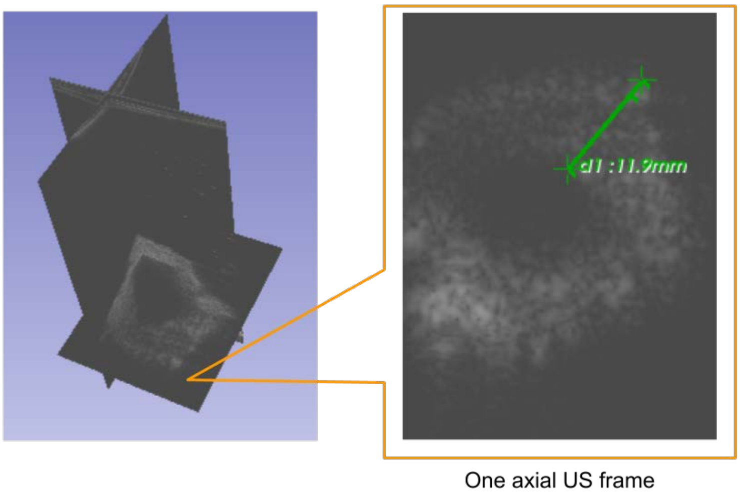


Figure 4.11 – Measuring the maximal and minimal distances between the ’tumor’ edge and the specimen edge on each axial US images by 3D Slicer software [Fedorov et al., 2012].

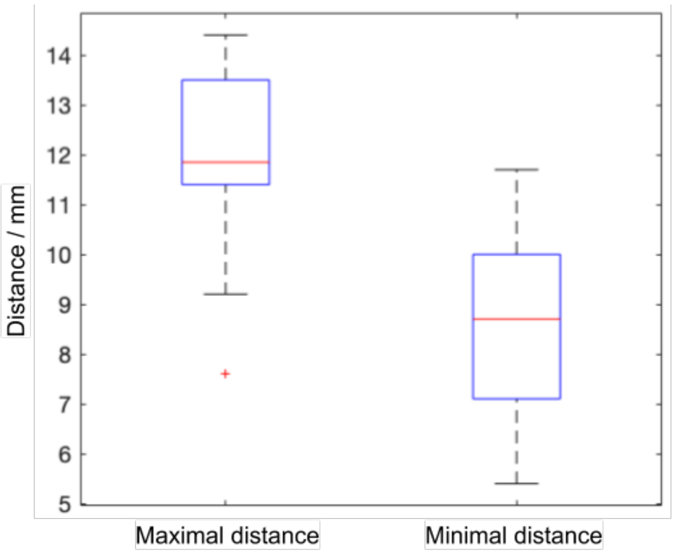


Figure 4.12 – Boxplot of the maximal and minimal distances between the ’tumor’ edge and the specimen edge on 10 axial US images.

Experimental results show that the implemented framework is able to present the edge of the simulated tumor to the user, and the resection error is smaller than 2 mm. Since we already evaluated the framework with rigid phantom previously in section 3.3 *Framework evaluation* and presented a millimetric error, this 2 mm error is mainly from the resection procedure. The resection skill depends on a surgeon’s experience. Accordingly, augmented endoscopic views should facilitate the tumor resection for novice surgeons, which deserves to be evaluated in a dedicated study.

4.3.3 Discussion of the ex-vivo lamb tongue experiment

As far as we know, this is the first proof-of-concept study that estimates the value of an intraoperative US imaging and AR technique in base of tongue cancer surgery.

The main contribution of this work is the efficient implementation of a 3D US-based AR framework. The proposed fully-automated 3D US probe calibration method does not require specific skills or user experience, and it takes less than 5 minutes. This fast and automatic calibration process increases the intraoperative use of 3D US imaging technique. Operating the endoscope in the working space of 20 mm to 60 mm from and $> 45°$ or perpendicular ($90°$) to the tongue base surface, the framework provides a high level of accuracy ($D_{rms} <$ 0.87 mm presented in section 3.3 *Framework evaluation*) for augmenting the endoscopic views. Each step of the framework was evaluated and showed high accuracy, as shown in the previous chapter .

In this experiment, we showed the information as an uniform green surface over the tongue view. Our framework can be used to find the best way to propose the information to surgeons. For example, as future work, we will explore two different way to display the information in the augmented view:

- The first could be more realistic, that is to display the polygon mesh of the tumor in the AR views, as shown in Figure 4.13. For this, we have to segment (manually for the moment) the tumors in US images. This segmentation is used to generate the 3D mesh model by surface modeling (*e.g.* provided by 3D Slicer software [Fedorov et al., 2012]).
- Another could be more informative, that is to augment the endoscopic views by projecting only tumor boundaries and resection margins (Figure 4.14). The surgeon needs only to delineate some tumor edges in the US volume. Then, the resection margins can be computed according to the tumor edge, and projected onto the tongue surface. Such a strategy has already be used in the experiment of a rectum phantom (section 4.4 *Validating the framework on a rectum phantom*). This solution not only

gives surgeons confidence in the reliability of tumor locations in US images, but also avoids the tedious and time-consuming manual US segmentation of the tumors.

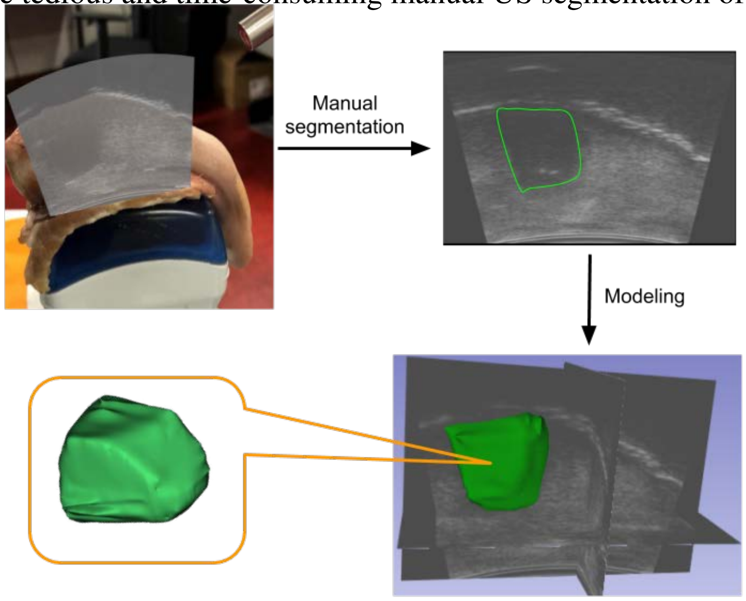
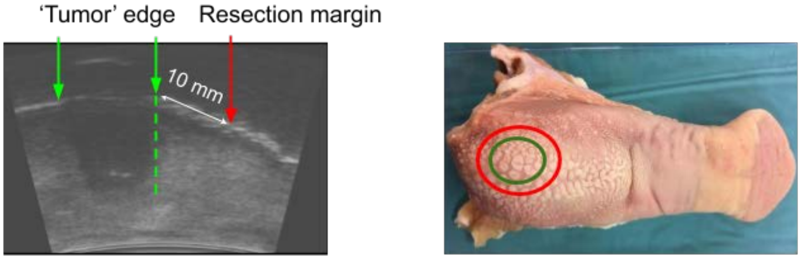


Figure 4.13 – Manually delineating ’tumor’ area on US image and surface modeling to create mesh model.



US image delineation

Augmenting with ‘tumor’ edge and resection margin

Figure 4.14 – Representation of ’tumor’ edge delineation (green) and resection margin generation (red) on an US image.

This experiment demonstrated the feasibility of applying the proposed framework in transoral robotic surgery for base of tongue cancer. Future work will validate this work on the Raven surgical research platform. Since the surgical treatment is ideal for treating the T1 and T2 base of tongue tumors which $<$ 4 mm, future work should include simulating small tumors in ex-vivo lamb tongue and evaluating the framework on such small tumors.

4.4 Validating the framework on a rectum phantom

4.4.1 Experimental setup

To integrate the AR framework into robot-assisted laparoscopic rectal surgery, we propose, with our clinical partners, the following surgical workflow after pelvic excision (*i.e.* before distal resection):

(1) A surgeon performs transrectal ultrasonography (TRUS) imaging on a patient's rectum

(2) The surgeon manually localizes the tumor edge on a 3D US image

(3) The implemented framework generates a distal resection margin (line)

(4) The framework transforms to, and superimposing on, the intraoperative endoscopic views with the tumor edge and the distal resection margin

(5) The surgeon marks the resection margin on the rectal wall by slightly cauterizing the tissues with an electrosurgical device, then, withdraws the TRUS probe from the patient.

(6) The surgeon performs distal rectal resection at the marked location

Based on above surgical workflow, an experimental validation setup is designed to simulate the surgical scenario. As shown in Figure 4.15, the rectum phantom is circular shape and made of dark brown silicone in which a simulated tumor is buried. This simulated tumor (circular shape of $r = 10\,\text{mm}$ and $h = 10\,\text{mm}$) is made of red and stiffer silicone and is not visible from the surface of the rectum phantom. The rectum phantom is designed to make the experiment as realistic as possible; in real surgery conditions, the tumor is usually concealed in the rectum and not visible in the endoscopic camera. In our experiment, the phantom was covered with a white paper, that allows the user to make a clear marking with a pen. After implementing the AR framework, the augmented view was displayed on the head-mounted display (HMD). A colorectal surgeon wore the HMD and, based on the augmented view, marked the proximal, distal and resection margins on the rectum phantom using a pen. This step was designed to simulate the slightly cauterizing of rectal tissues with an electrosurgical device during surgery. In rectal surgery, the surgeon needs to mark only the resection margin which corresponds to the stapling location. Nevertheless, in order to evaluate the performance of the proposed AR framework, during the experiment, the surgeon was asked to mark the proximal and distal end of the simulated tumor (Figure 4.17 (a)). Finally, the TRUS probe was withdrawn from the rectum phantom and the AR framework was stopped. The phantom was cut following the marked lines and the cutting planes were analyzed, as shown in Figure 4.17 (b).

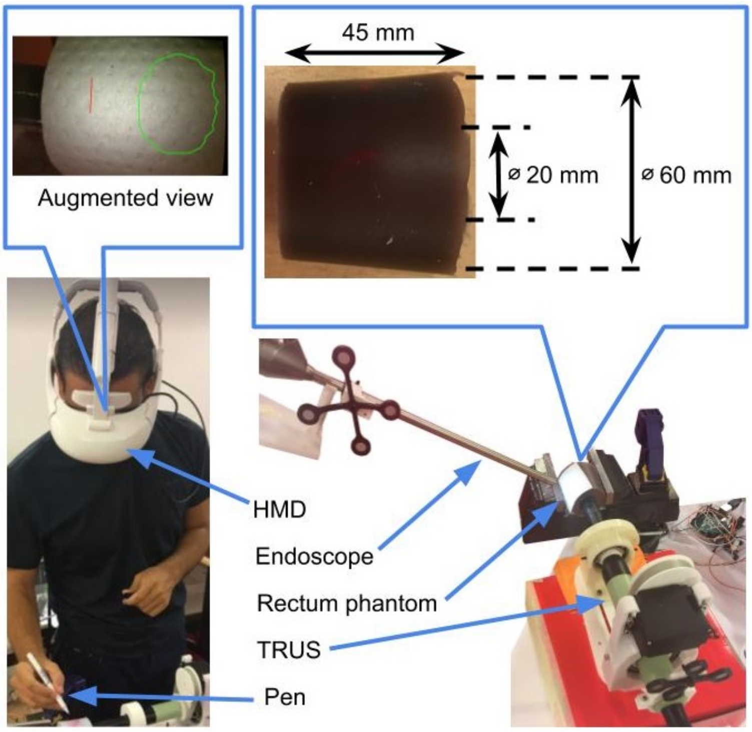


Figure 4.15 – Experiment design: surgeon localizing the simulated rectal tumor based on the augmented endoscopic view.

4.4.2 Endoscopic image augmentation

Figure 4.16 shows the process of endoscopic image augmentation using the virtual information generated from a 3D US image. The rectum phantom is imaged by the TRUS probe. 3D US reconstruction is performed with k-Wave MATLAB® toolbox [Treeby and Cox, 2010]. The augmentation steps in Figure 4.16 are:

(1) manually annotating the distal and proximal margins of the simulated tumor on the 'rectal wall' in 20 US frames (this manual annotation of these 40 points took 1 to 2 min for our surgeon);

(2) automatically finding the most distal margin in the 3D US image and generating the resection margin (line) 10 mm below the most distal margin;

(3) collecting the locations of the annotations from 20 US frames;

(4) projecting the coordinates of the annotations to the endoscopic view based on the registration result between the US image and the endoscopic camera (as shown in equation 3.7).

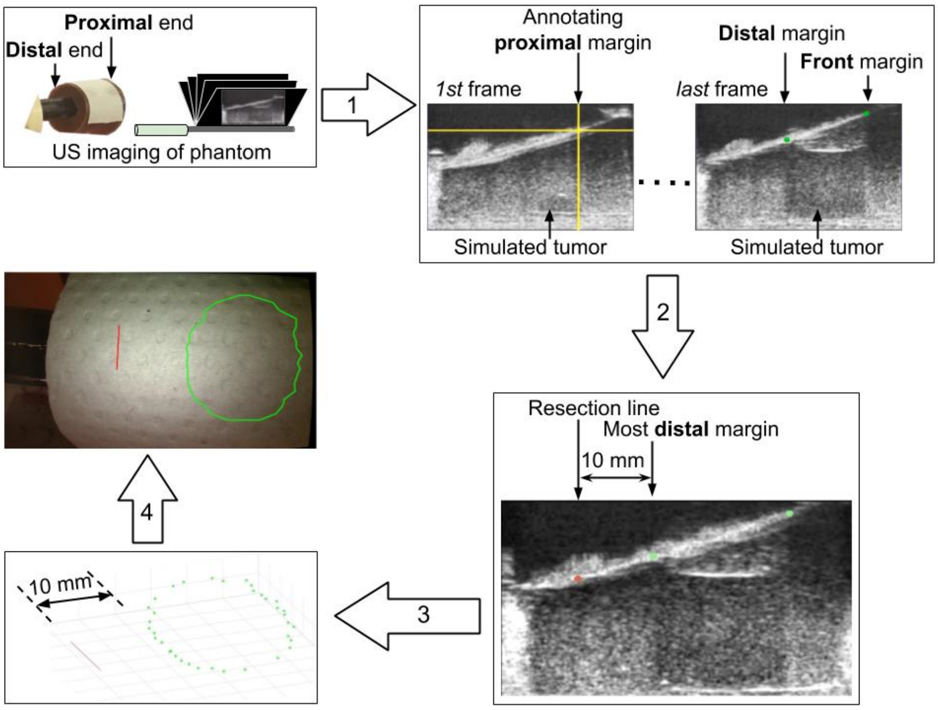


Figure 4.16 – Endoscopic image augmentation: (1) annotating distal and proximal margin (green) of the simulated tumor on 'rectal wall' in US images using crosshairs (yellow); (2) generating resection line (red) 10 mm below the most distal margin; (3) collecting the locations of these annotations; (4) projecting this information over the endoscopic view.

4.4.3 Results of the rectal phantom experiment

During the experiment, the surgeon marked the proximal and distal end of the simulated tumor and the distal resection margin on the rectum phantom by using the implemented AR framework, as shown in Figure 4.17 (a). Then, the surgeon made straight cuts on these marks (Figure 4.17 (a)) through the phantom and analyzed the cutting surfaces (Figure 4.17 (b)), in order to examine whether the marks of the proximal/distal margins are the real proximal/distal ends of the simulated tumor or not.

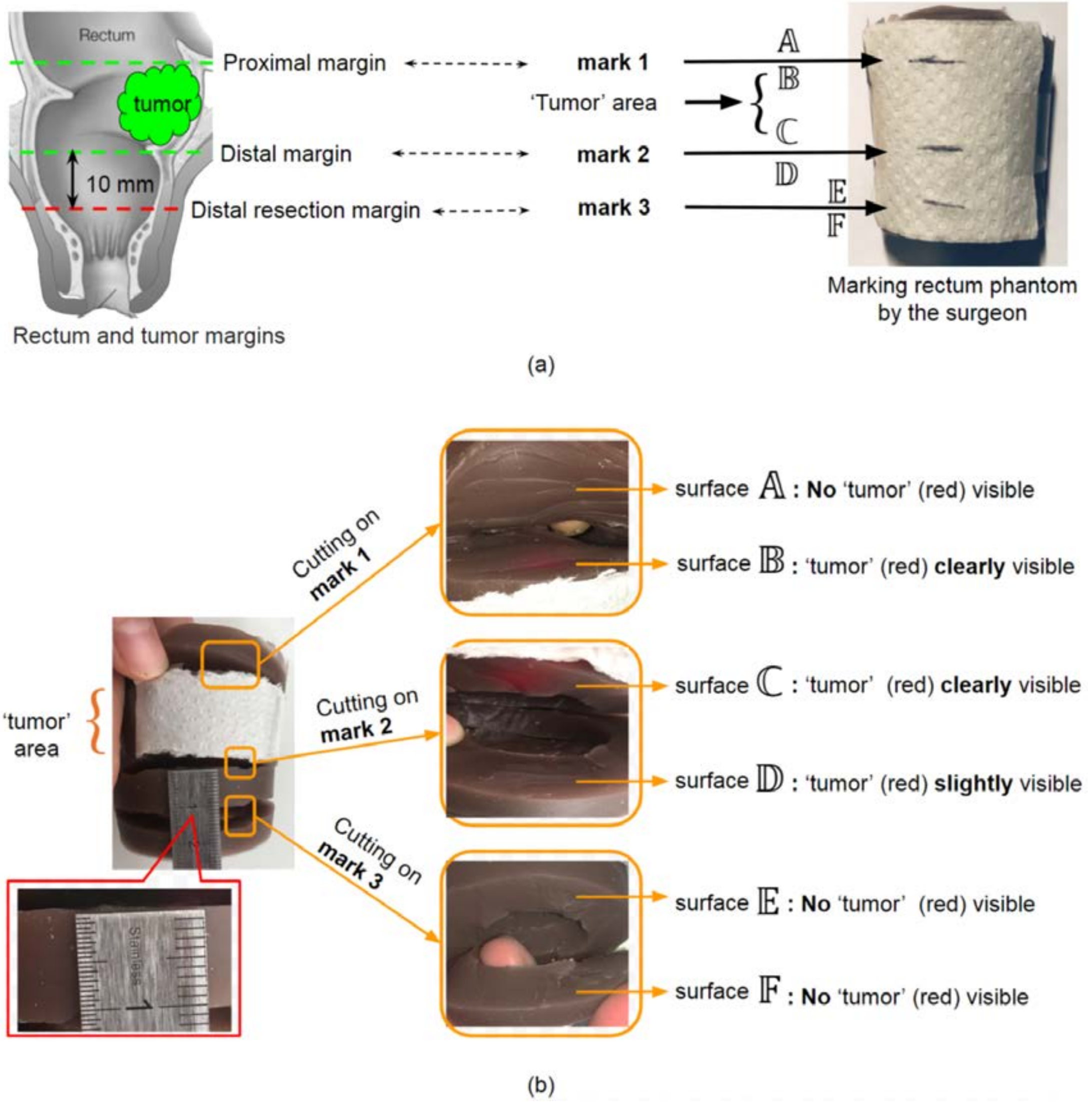
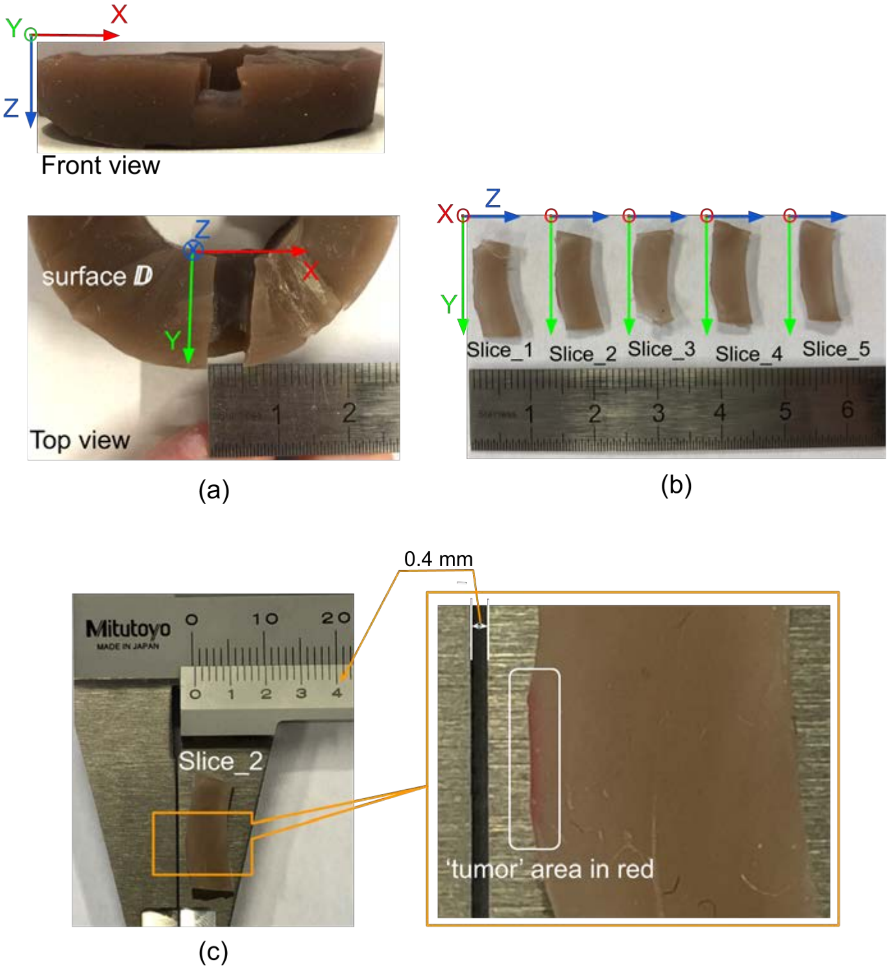


Figure 4.17 – Experimental results: (a) markings of the proximal, distal and resection margins by a pen; (b) cutting planes $\mathbb{A}$ to $\mathbb{F}$ analysis.

As shown in Figure 4.17 (b), the simulated tumor (red) is not seen on surfaces $\mathbb{A}$ but clearly appears on surface $\mathbb{B}$. That indicates **mark 1** is on the proximal margins of the simulated tumor. The distance between **mark 2** and **mark 3** was measured by a ruler giving 10 mm spacing (Fig. 4.17 (b)). This validates that **mark 3** is on the distal resection margin. As shown in Fig. 4.17 (b), the red 'tumor' area is clearly seen on surface $\mathbb{C}$ but appears on surface $\mathbb{D}$ (3 mm × 3 mm red area) unexpectedly. In order to verify the accuracy of the estimated distal end of the 'tumor' (i.e. **mark 2**), we resected this 'tumor' area into 5 longitudinal slices (approximately 1 mm per slice, as shown in Fig. 4.18 (a) and (b)). In the worst case, the red ’tumor’ area appeared at a depth < 0.4 mm in surface $\mathbb{D}$ (Fig. 4.18 (c)).

Experimental results show that the implemented framework is able to accurately localize the edge of the simulated tumor and to display the 10 mm distal resection margin to the surgeon, otherwise determined on the basis of the surgeon’s experience. Accordingly, this augmented view of the distal resection margin should facilitate the resection step for novice surgeons. This particular point deserves to be evaluated in a dedicated study.



(a)

(b)

(c)

Figure 4.18 – Analyzing how deep the ’tumor’ infiltrating into surface $\mathbb{D}$ using a vernier scale.

4.4.4 Discussion of the rectal phantom experiment

To the best of our knowledge, this is the first proof-of-concept study that estimates the value of an intraoperative TRUS imaging and AR technique in laparoscopic rectal cancer surgery.

TRUS is chosen as an intraoperative imaging modality to image rectal cancer and provide virtual information to augment the surgeon's view. US image is commonly used in rectal cancer staging and compares favorably to other intraoperative imaging modalities (e.g. cone beam computed tomography or open MRI). Indeed, US systems provide real-time imaging; they are less expensive, non-ionizing and easier to integrate in the operating room. In this work, a fast and accurate phantom-based US calibration method was developed to increase the use of the 3D US imaging technique in the operating room. Our fully-automated calibration procedure requires no specific skills or user experience. It takes approximately 5 minutes to calibrate a 3D US probe. Furthermore, this method works on both the motorized TRUS probe and the built-in array 3D TRUS probe, thereby allowing the integration of both types of 3D probes in our framework.

The hand-eye calibration method insures high accuracy ($D_{rms} < 0.51$ mm) when the camera is operated 20 mm to 60 mm from and $> 45^\circ$ or perpendicular (90°) to the surface of the target object. When the endoscopic camera is positioned in extreme poses, the calibration error remains low ($D_{rms} < 1.52$ mm) compared to the resection margin position (10 mm to the distal end of the tumor). In robot-assisted laparoscopic rectal surgery, before distal resection, the endoscopic camera is positioned in front of the rectum, as shown in Figure 4.19. This is an optimal pose in which the camera has been calibrated to provide the best accuracy < 0.51 mm. During this experiment, the hand-eye calibration takes approximately 10 minutes. This step can be executed during patient installation and should not disrupt the surgical workflow.

For clinical application purposes, we proposed with our clinical partners the following adaptations:

— **Intraoperative TRUS imaging of rectum** For the clinical implementation of our framework, we plan to use 3D TRUS probes (e.g. BK Medical® Endocavity 3D 8838) with built-in linear array rotating inside of the probe to ensure stable contact

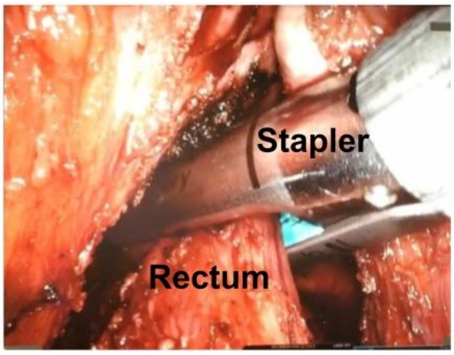


Figure 4.19 – An endoscopic view of distal resection of rectum in robot-assisted laparoscopic surgery

(no relative movement) of the rectum and the probe. The proposed phantom-based US probe calibration method is also suitable for calibrating this kind of 3D TRUS probe [Shen et al., 2018].

- **Semi-automatic margin generation** As presented in section 1.2.1 *Surgical treatment for low-rectal cancer*, the distal resection of the rectum is one of the most challenging tasks for colorectal surgeons. Therefore, during surgery, only the distal resection margin is important to show to colorectal surgeons. In order to quickly determine this margin, we developed a semi-automatic method:

  (1) A surgeon manually selects several US frames in which the tumor is located relatively lower than that in other US frames.

  (2) The surgeon manually marks the distal end of the tumor on the selected US frames.

  (3) The coordinates of these markings in the 3D US image are automatically compared to each other and the lowest coordinate is selected to generate the distal margin of the tumor.

  (4) The resection margin is automatically generated 10 mm below the distal margin of the tumor.

  This semi-automatic method avoids the tedious and time-consuming manual US segmentation of the tumor. In our experiment, it took less than 1 minute to perform the manual annotation on the rectum phantom. However, this step must, and will be, evaluated with surgeons to validate the maximum acceptable working time of this semi-automatic method during actual surgery.
- **Marking margins on rectum** Our framework does tackle the challenge of showing

the distal resection margin to surgeons. However, maintaining the accurate augmented view during surgery requires the TRUS probe to remain operational in the rectum, that prevents the rectal resection. In order to solve this problem, we propose to keep the probe in the rectum and use an electrosurgical device to slightly cauterize the rectal tissue on the virtual resection margin. These cauterized marks help surgeons to recognize the resection margin, after withdrawing the TRUS probe. Such electrosurgical devices are widely accessible in operating rooms. For instance, the da Vinci® surgical platform features a power generator for electrosurgical instruments.

4.5 Conclusion

In robot-assisted surgery, successful surgical outcomes depend on two main factors: surgical guidance (the purpose of our AR framework), and the surgeon's robotic surgical training. Since the latter is beyond the scope of this thesis, in our experiments, the surgeon performed resection without a surgical robot.

In this chapter, we validate the framework on an ex-vivo lamb tongue and a rectal phantom and show the possibility of integrating the implemented AR guidance into surgical workflow. Future work will validate the proposed framework on the *Raven* surgical research platform with an ex-vivo tongue or a rectum. In order to quantify the added value of this framework in terms of improvement of the surgical learning curve, the surgical gesture precision of novice and expert surgeons should be compared in the future.

Conclusion and outlooks

This thesis introduces robot-assisted surgery for treating base of tongue cancer and low-rectal cancer, based on which we have implemented an augmented reality (AR) framework for tumor resection guidance.

Tumor resection is a particularly important step in surgery, because incomplete resection causes a high risk of cancer recurrence [Vermaas et al., 2007]. However, during robot-assisted minimally invasive surgery, it is a challenge for surgeons to identify clear resection margins of base of tongue tumors or low-rectal tumors, because of following reasons:

(1) Tumors are often concealed in organs and not visible in intraoperative endoscopic views.

(2) Manual palpation to diagnose an operative field is difficult to perform during robot-assisted tumor resection. For example, da Vinci® surgical instruments (*i.e.* an endoscope, a stapler and a gripper) operating on a patient's base of tongue prevent surgeons from inserting fingers to the patient's oral cavity for palpation.

(3) The present robotic surgical instruments, for instance, da Vinci® surgical instruments do not provide tactile feedback, thereby preventing surgeons from telling the difference between dissecting pathologic tissues and dissecting normal tissues.

(4) During image-guided procedures, surgeons have to mentally deform patient-specific medical images to the operative fields to make resection plans. The accuracy of tumor resection is highly dependent upon the surgeons' experience.

Intraoperative guidance systems have been proposed for a number of surgical procedures, such as laparoscopic liver surgery [Feuerstein et al., 2008] and image-guided prostatectomy[Ukimura et al., 2010]. The development of such a system is based on the surgical workflow. Unfortunately, to the best of our knowledge, intraoperative AR guidance has not been developed for robot-assisted laparoscopic rectal surgery. Moreover, for transoral robotic surgery, only few studies have implemented AR guidance systems. However, these studies either do not take into account intraoperative tongue deformation properly (which uses preoperative images to localize tumors) [Pratt and Arora, 2018] or use ionizing radiation-based intraoperative imaging technique (*i.e.* cone beam computed tomography (CBCT) and fluoroscopy) [Liu et al., 2015]. Literature has shown that tumors in base of tongues and rectums can be visualized by the diagnostic imaging technique — ultrasonography (US), as presented in Chapter 1 *Medical context and image-guidance systems*, which motivates us to develop an intraoperative US-based AR framework for tumor resection guidance in robot-assisted base of tongue cancer (or low-rectal cancer) surgery.

In order to integrate our AR framework into the surgical workflow, we proposed following steps with our clinical partner:

(1) In surgery, before the step of tumor resection, a surgeon perform US imaging on the submental area of a patient to image the base of tongue tumor (or perform transrectal US imaging on a patient's rectum to image the low-rectal tumor).

(2) The intraoperative US image is manually segmented to extract the tumor area, which can be used to create a virtual model for augmenting the endoscopic views.

(3) Based on optical tracking technique and an accurate registration of the US image to the endoscopic image, the proposed AR framework transformed the tumor model to the intraoperative endoscopic views to display. Moreover, the distal resection margin for rectum resection is also displayed on the endoscopic views.

(4) According to the augmented endoscopic views, the surgeon can mark the tumor area or resection margins by using an electrosurgical device to slightly cauterize the surface of the base of tongue (or the rectum wall). This is to prevent the intraoperative tissue deformation from disturbing the AR accuracy.

(5) The surgeon withdraws the US probe from the patient and resects the tumor following the markings.

The accuracy of the AR guidance framework mainly depends on the registration of the US image to the endoscopic image. To provide an accurate AR guidance, we used optical tracking technique, which is a favorable tracking technique with high accuracy and stability [Birkfellner et al., 2008; Engelhardt et al., 2016]. Two markers were respectively attached to the US probe and the endoscopic camera for tracking. For calibrating the US image with respect to the coordinate system of the marker of the US probe, we developed an automatic and accurate 3D US probe calibration method based on a 3D-printed phantom [Shen et al., 2018]. We implemented popular hand-eye calibration methods [Tsai and Lenz, 1989; Daniilidis, 1999; Malti and Barreto, 2010] for calibrating the endoscopic views with respect to the coordinate system of the marker of the endoscopic camera. The optical tracking technique and these two accurate calibration processes ensure the proper registration of the US images to the endoscopic views. The implemented AR framework was evaluated by projecting the US segmentation of a calibration phantom onto endoscopic views. It shows the error between the projected information and the calibration phantom in the endoscopic views less than 1 mm. Moreover, our AR framework has been validated on three silicone phantoms, a rectum phantom and an ex-vivo lamb tongue. During the experimental validation, the AR framework was implemented to guide the participants and a colorectal surgeon to resect simulated tumors in those phantoms. These experiments are designed to simulate integrating our framework into the surgical workflow of robot-assisted base of tongue (and rectal) surgery. The experimental results show that, according to the augmented views provided by the proposed AR framework, the participants and the colorectal surgeon are able to accurately resect the simulated tumors.

To the best of our knowledge, it is the first study to show the feasibility of using intraoperative US technique and AR technique to guide tumor resection in transoral robotic surgery for base of tongue cancer and robot-assisted laparoscopic low-rectal surgery. Moreover, we developed an automatic US probe calibration method which streamlines the calibration process in literature to a registration of the 3D US image to the CAD model of a 3D-printed calibration phantom. Furthermore, the proposed calibration method is easy to implement and use, and provides a high level of accuracy. It facilitates the intraoperative use of the US technique. Finally, the experimental validation is carried out based on two specific surgical procedures (*i.e.* robot-assisted surgery for base of tongue cancer and low-rectal cancer) and shows the feasibility of integrating our AR framework into the surgical workflow.

Future work of this project will mainly focus on developing an US image segmentation method for segmenting the base of tongue cancer and rectal cancer. Besides, the virtual model of the tumor should display on the endoscopic view using polygon mesh, so that the augmented view will show the depth of tumor invasion. In order to quantify the added value of the implemented framework in terms of improvement of the surgical learning curve, the surgical gesture precision of novice and expert surgeons should be evaluated in a dedicated study. Besides, future work will validate the proposed framework on the Raven surgical research platform.

Outlooks

Over the past decades, minimally invasive surgery has been popular due to the clinical benefits. However, surgeons lose the direct vision and tactical feedback on the operative field during minimally invasive interventions. Integrating AR technique into the surgical workflow can be a solution to alleviate this drawback. For the guidance of tumor resection in soft tissue surgery, the use of intraoperative US and AR techniques is promising: US images can show the information of soft-tissue tumors (*e.g.* tumor location and shape) and AR technique can display this information onto surgical scenes (*i.e.* intraoperative endoscopic views).

The real-time imaging technique — intraoperative US allows surgeons to adjust patient-specific planning intra-operatively to adapt different surgical strategies. Moreover, a number of manufacturers, such as Clarius Mobile Health, Siemens Healthineers, Philips, Sonoscanner and so on, have developed advanced wireless portable US systems, that significantly facilitates the intraoperative use of US imaging technique. In order to obtain complementary information of the operative field, multi-modality images (such as intraoperative US images and preoperative MRI/CT images) registration and fusion become increasingly important. A large number of studies have proposed US-and-MRI fusion techniques for navigation in soft tissue surgery, such as liver surgery [Porter et al., 2001; Lange et al., 2003]. These fusion techniques have been also used for tongue contour tracking (mostly for the analyses of the human tongue motion) [Lulich, 2018]. In the future, the development of US-and-MRI fusion technique can provide more complementary information of base of tongue cancer (or rectal cancer) for surgical AR guidance.

Currently, the registration of the intraoperative US images to the surgical scenes remains a challenging task in soft tissue surgery, due to the non-rigidity of tissues and organs, and the limitations of present tracking techniques. For instance, optical tracking techniques have restriction of line-of-sight requirement [Ukimura et al., 2010], electromagnetic tracking techniques provide small working volume and suffer from electromagnetic interference [Franz et al., 2014], computer-vision-based tracking techniques highly rely on the pattern and the condition of the used markers [Singla et al., 2017]. Recently, some studies used photoacoustic (PA) markers as registration landmarks [Cheng et al., 2014]. These PA markers can be generated intra-operatively, however, the main disadvantage is the safety concern: a laser delivery system guides the laser beam into a patient's body and fire it on the surface of an organ. Overall, the development of new tracking techniques is important for intraoperative US-based AR systems.

In transoral robotic surgery for base of tongue tumor resection, it is important for surgeons to assess the tumor invasion depth. However, one of primary issues of AR visualization is to provide correct depth perception which is disturbed by the overlay of real and virtual information [Bichlmeier et al., 2007]. Some studies proposed to display the distance between the tip of the surgical tool and the nearest organ in the augmented surgeons' views [Choi et al., 2016] or project a depth bar to show the depth of needle insertion [Gavaghan et al., 2012]. Moreover, advanced biomechanical models of organs [Haouchine et al., 2013; Demongeot et al., 2016] can be used in AR systems to display the tumor deformation following the motion of the operative field, when the surgical instruments operate on it. These techniques can be used to show to surgeons the proper resection depth in AR-guided transoral robotic surgery.

This thesis presents two surgical procedures (*i.e.* transoral robotic surgery for base of tongue cancer and robot-assisted laparoscopic low-rectal surgery), based on which we implemented an intraoperative US-based AR framework for guiding tumor resection. The framework is evaluated and provides a high-level of accuracy. The experimental validation shows the feasibility of integrating the proposed framework into surgical workflow.

Bibliography

Abeysekera, J. M., Najafi, M., Rohling, R., and Salcudean, S. E. (2014). Calibration for Position Tracking of Swept Motor 3-D Ultrasound. *Ultrasound in Medicine and Biology*, 40(6):1356–1371.

Agustinos, A., Wolf, R., Long, J. A., Cinquin, P., and Voros, S. (2014). Visual servoing of a robotic endoscope holder based on surgical instrument tracking. In *IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics*, pages 13–18.

Atallah, S., Larach, S. W., and Monson, J. R. (2016). Stereotactic navigation for TAMIS-TME. *Minimally Invasive Therapy and Allied Technologies*, 25(5):271–277.

Azuma, R. T. (1997). A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385.

Baek, S. J., Kim, C. H., Cho, M. S., Bae, S. U., Hur, H., Min, B. S., Baik, S. H., Lee, K. Y., and Kim, N. K. (2015). Robotic surgery for rectal cancer can overcome difficulties associated with pelvic anatomy. *Surgical Endoscopy*, 29(6):1419–1424.

Bergmeir, C., Seitel, M., Frank, C., De Simone, R., Meinzer, H.-P., and Wolf, I. (2009). Comparing calibration approaches for 3D ultrasound probes. *International Journal of Computer Assisted Radiology and Surgery*, 4:203–213.

Bernhardt, S., Nicolau, S. A., Soler, L., and Doignon, C. (2017). The status of augmented reality in laparoscopic surgery as of 2016. *Medical Image Analysis*, 37:66–90.

Bichlmeier, C., Wimmer, F., Heining, S. M., and Navab, N. (2007). Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality. In *6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 129–138.

Birkfellner, W., Hummel, J., and Wilson, E. (2008). Tracking devices. In Peters, T. and Cleary, K., editors, *Image-Guided Interventions*, number 2008, chapter 2, pages 23–45. Springer, Boston, MA, Boston, MA.

Blanco, R. G. F., Califano, J., Messing, B., Richmon, J., Liu, J., Quon, H., Neuner, G., Saunders, J., Ha, P. K., Sheth, S., Gillison, M., Fakhry, C., Gervacio, R., Blanco, F., Califano, J., Messing, B., Richmon, J., Liu, J., Quon, H., Neuner, G., Saunders, J., Ha, P. K., Sheth, S., Gillison, M., and Fakhry, C. (2014). Transcervical ultrasonography is feasible to visualize and evaluate base of tongue cancers. *PLoS ONE*, 9(1):e87565.

Branstetter, B. F., Blodgett, T. M., Zimmer, L. A., Snyderman, C. H., Johnson, J. T., Raman, S., and Meltzer, C. C. (2005). Head and neck malignancy: is PET/CT more accurate than PET or CT alone? *Radiology*, 235(2):580–586.

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 68(6):394–424.

Chen, X., Xu, L., Wang, Y., Wang, H., Wang, F., Zeng, X., Wang, Q., and Egger, J. (2015). Development of a surgical navigation system based on augmented reality using an optical see-through head-mounted display. *Journal of Biomedical Informatics*, 55:124–131.

Cheng, A., Guo, X., Kang, H.-J., Tavakoli, B., Kang, J. U., Taylor, R. H., and Boctor, E. M. (2014). Concurrent photoacoustic markers for direct three-dimensional ultrasound to video registration. In Oraevsky, A. A. and Wang, L. V., editors, *SPIE 8943, Photons Plus Ultrasound: Imaging and Sensing 2014*, volume 8943, page 89435J.

Cheung, C. L., Wedlake, C., Moore, J., Pautler, S. E., and Peters, T. M. (2010). Fused video and ultrasound images for minimally invasive partial nephrectomy: A phantom study. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 408–415.

Choi, H., Cho, B., Masamune, K., Hashizume, M., and Hong, J. (2016). An effective visualization technique for depth perception in augmented reality-based surgical navigation. *International Journal of Medical Robotics and Computer Assisted Surgery*, 12(1):62–72.

Clayburgh, D. R., Byrd, J. K., Bonfili, J., and Duvvuri, U. (2016). Intraoperative Ultrasonography during Transoral Robotic Surgery. *Annals of Otology, Rhinology and Laryngology*, 125(1):37–42.

Comeau, R. M., Sadikot, A. F., Fenster, A., and Peters, T. M. (2000). Intraoperative ultrasound for guidance and tissue shift correction in image-guided neurosurgery. *Medical Physics*, 27(4):787–800.

Coquia, S. F., Hamper, U. M., Holman, M. E., DeJong, M. R., Subramaniam, R. M., Aygun, N., and Fakhry, C. (2015). Visualization of the oropharynx with transcervical ultrasound. *American Journal of Roentgenology*, 205(6):1288–1294.

Daniilidis, K. (1999). Hand-eye calibration using dual quaternions. *The International Journal of Robotics Research*, 18(3):286–298.

Daniilidis, K. and Bayro-Corrochano, E. (1996). The dual quaternion approach to hand-eye calibration. *Proceedings - International Conference on Pattern Recognition*, 1:318–322.

Demongeot, J., Lontos, A., and Promayon, E. (2016). Discrete Mesh Approach in Morphogenesis Modelling: the Example of Gastrulation. *Acta Biotheoretica*, 64(4):427–446.

Desai, S. C., Sung, C.-K., and Genden, E. M. (2008). Transoral robotic surgery using an image guidance system. *The Laryngoscope*, 118(11):2003–2005.

Detmer, P. R., Bashein, G., Hodges, T., Beach, K. W., Filer, E. P., Burns, D. H., and Strandness, D. E. (1994). 3D ultrasonic image feature localization based on magnetic scanhead tracking: in vitro calibration and validation. *Ultrasound in Medicine and Biology*, 20(9):923–936.

Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302.

Dubuisson, M. and Jain, A. K. (1994). A modified Hausdorff distance for object matching. In *12th International Conference on Pattern Recognition*, pages 566–568.

Engelhardt, S., De Simone, R., Al-Maisary, S., Kolb, S., Karck, M., Meinzer, H. P., and Wolf, I. (2016). Accuracy evaluation of a mitral valve surgery assistance system based on optical tracking. *International Journal of Computer Assisted Radiology and Surgery*, 11(10):1891–1904.

Fakhry, C., Agrawal, N., Califano, J., Messing, B., Liu, J., Saunders, J., Ha, P., Coquia, S., Hamper, U., Gillison, M., and Blanco, R. (2014). The use of ultrasound in the search for the primary site of unknown primary head and neck squamous cell cancers. *Oral Oncology*, 50(7):640–645.

Farwell, D. G., Phipps, J., Gorpas, D., Ma, D., Bec, J., Yankelevich, D., Sorger, J., Gandour-Edwards, R., Bewley, A., and Marcu, L. (2016). Augmented fluorescence lifetime imaging for enhanced tumor margin delineation during trans-oral robotic surgery. In *Proceedings of the 9th International Conference on Head and Neck Cancer*.

Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J. C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., Buatti, J., Aylward, S., Miller, J. V., Pieper, S., and Kikinis, R. (2012). 3D Slicer as an image computing platform for the quantitative imaging network. *Magnetic Resonance Imaging*, 30(9):1323–1341.

Feuerstein, M., Mussack, T., Heining, S., and Navab, N. (2008). Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection. *IEEE Transactions on Medical Imaging*, 27(3):355–369.

Feuerstein, M., Reichl, T., Vogel, J., Traub, J., and Navab, N. (2009). Magneto-optical tracking of flexible laparoscopic ultrasound: model-based online detection and correction of magnetic tracking errors. *IEEE Transactions on Medical Imaging*, 28(6):951–967.

Fichtinger, G., Deguet, A., Masamune, K., Balogh, E., and Fischer, G. (2005). Image overlay guidance for needle insertion in CT scanner. *IEEE Transactions on Biomedical Engineering*, 52(8):1415–1424.

Fitzpatrick, J. M., West, J. B., and Maurer, C. R. (1998). Predicting error in rigid-body point-based registration. *IEEE Transactions on Medical Imaging*, 17(5):694–702.

Franz, A. M., Haidegger, T., Birkfellner, W., Cleary, K., Peters, T. M., and Maier-Hein, L. (2014). Electromagnetic tracking in medicine -A review of technology, validation, and applications. *IEEE Transactions on Medical Imaging*, 33(8):1702–1725.

Gavaghan, K., Oliveira-Santos, T., Peterhans, M., Reyes, M., Kim, H., Anderegg, S., and Weber, S. (2012). Evaluation of a portable image overlay projector for the visualisation of surgical navigation data: Phantom studies. *International Journal of Computer Assisted Radiology and Surgery*, 7(4):547–556.

Gerig, G. and Klein, F. (1986). Fast contour identification through efficient hough transform and simplified interpretation strategy. In *International Conference on Pattern Recognition*, pages 498–500. IEEE.

Greif, F., Aranovich, D., Hananel, N., Knizhnik, M., and Belenky, A. (2009). Intraoperative ultrasound in colorectal surgery. *Journal of Clinical Ultrasound*, 37(7):375–379.

Haouchine, N., Dequidt, J., Peterlik, I., Kerrien, E., Berger, M.-O., and Cotin, S. (2013). Image-guided simulation of heterogeneous tissue deformation for augmented reality during hepatic surgery. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 199–208.

Hsu, P.-W., Prager, R. W., Houghton, N. E., Gee, A. H., and Treece, G. M. (2007). Accurate fiducial location for freehand 3D ultrasound calibration. In Emelianov, S. Y. and McAleavey, S. A., editors, *SPIE 6513, Medical Imaging 2007: Ultrasonic Imaging and Signal Processing*, volume 6513, page 651315. Society of Photo-Optical Instrumentation Engineers (SPIE).

Hu, Y., Ahmed, H. U., Taylor, Z., Allen, C., Emberton, M., Hawkes, D., and Barratt, D. (2012). MR to ultrasound registration for image-guided prostate interventions. *Medical Image Analysis*, 16(3):687–703.

Huh, C. H., Bhutani, M. S., Farfán, E. B., and Bolch, W. E. (2003). Individual variations in mucosa and total wall thickness in the stomach and rectum assessed via endoscopic ultrasound. *Physiological measurement*, 24(4):N15–22.

Kenngott, H. G., Wagner, M., Gondan, M., Nickel, F., Nolden, M., Fetzer, A., Weitz, J., Fischer, L., Speidel, S., Meinzer, H. P., Böckler, D., Büchler, M. W., and Müller-Stich, B. P. (2014). Real-time image guidance in laparoscopic liver surgery: First clinical experience with a guidance system based on intraoperative CT imaging. *Surgical Endoscopy*, 28(3):933–940.

Kim, N. K., Kim, M. J., Yun, L.-S. H., Sohn, S. K., and Min, J. S. (1999). Comparative study of transrectal ultrasonography, pelvic computerized tomography, and magnetic resonance imaging in preoperative staging of rectal cancer. *Diseases of the Colon and Rectum*, 42(6):770–775.

Kisansa, M. E. and Andronikou, S. (2017). Ultrasound imaging of tongue malignancy. *International Journal of Case Reports and Images*, 8(1):1–6.

Konishi, K., Hashizume, M., Nakamoto, M., Kakeji, Y., Yoshino, I., Taketomi, A., Sato, Y., Tamura, S., and Maehara, Y. (2005). Augmented reality navigation system for endoscopic surgery based on three-dimensional ultrasound and computed tomography: Application to 20 clinical cases. *International Congress Series*, 1281:537–542.

Kruskal, J. B. and Kane, R. A. (2006). Intraoperative US of the liver: techniques and cilnical applications. *RadioGraphics*, 26(4):1067–1084.

Lange, T., Eulenstein, S., Hiinerbein, M., and Schlag, P.-m. (2003). Vessel-based non-rigid registration of MR / CT and 3D ultrasound for navigation in liver surgery. *Computer Aided Surgery*, 8(5):228–240.

Liu, H. E., Auvinet, E., Giles, J., Rodriguez, F., and Baena, Y. (2018). Augmented reality based navigation for computer assisted hip resurfacing: a proof of concept study. *Annals of Biomedical Engineering*, 46(10):1595–1605.

Liu, W. P., Reaugamornrat, S., Deguet, A., Sorger, J. M., Siewerdsen, J. H., Richmon, J., and Taylor, R. H. (2013). Toward intraoperative image-guided transoral robotic surgery. *Journal of robotic surgery*, 7(3):217–225.

Liu, W. P., Richmon, J. D., Sorger, J. M., Azizian, M., and Taylor, R. H. (2015). Augmented reality and cone beam CT guidance for transoral robotic surgery. *Journal of Robotic Surgery*, 9(3):223–233.

Liu, X., Kang, S., Plishker, W., Zaki, G., Kane, T. D., and Shekhar, R. (2016). Laparoscopic stereoscopic augmented reality: toward a clinically viable electromagnetic tracking solution. *Journal of Medical Imaging*, 3(4):045001.

Lulich, S. M. (2018). Registration and fusion of 3D head-neck MRI and 3D/4D tongue ultrasound. *The Journal of the Acoustical Society of America*, 144(3):1904–1904.

Malti, A. and Barreto, J. P. (2010). Robust hand-eye calibration for computer aided medical endoscopy. In *2010 IEEE International Conference on Robotics and Automation*, pages 5543–5549, Alaska.

Marques, B., Roy, F., Haouchine, N., Jeanvoine, E., Cotin, S., Plantefeve, R., and Peterlik, I. (2015). Framework for augmented reality in Minimally Invasive laparoscopic surgery. In *17th International Conference on E-Health Networking, Application and Services (HealthCom), IEEE*, pages 22–27.

Masutani, Y., Dohi, T., Yamane, F., Iseki, H., and Takakura, K. (1998). Augmented reality visualization system for intravascular neurosurgery. *Computer Aided Surgery*, 3(5):239–247.

Mountney, P., Fallert, J., Nicolau, S., Soler, L., and Mewes, P. W. (2014). An augmented reality framework for soft tissue surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 423–431.

Nakamoto, M., Nakada, K., Sato, Y., Konishi, K., Hashizume, M., and Tamura, S. (2008). Intraoperative magnetic tracker calibration using a magneto-optic hybrid tracker for 3-D ultrasound-based navigation in laparoscopic surgery. *IEEE Transactions on Medical Imaging*, 27(2):255–270.

Nicolaou, M., James, A., Lo, B. P. L., Darzi, A., and Yang, G.-Z. (2005). Invisible shadow for navigation and planning in minimal invasive surgery. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 25–32.

Patsias, A., Giraldez-Rodriguez, L., Polydorides, A. D., Richards-Kortum, R., Anandasabapathy, S., Quang, T., Sikora, A. G., and Miles, B. (2015). Feasibility of transoral robotic-assisted high-resolution microendoscopic imaging of oropharyngeal squamous cell carcinoma. *Head & Neck*, 37(8):E99–E102.

Poon, T. C. and Rohling, R. N. (2005). Comparison of calibration methods for spatial tracking of a 3-D ultrasound probe. *Ultrasound in Medicine & Biology*, 31(8):1095–1108.

Porter, B. C., Rubens, D. J., Strang, J. G., Smith, J., Totterman, S., and Parker, K. J. (2001). Three-dimensional registration and fusion of ultrasound and MRI using major vessels as fiducial Markers. *IEEE Transactions on Medical Imaging*, 20(4):354–359.

Prager, R. W., Rohling, R. N., Gee, A. H., and Berman, L. (1998). Rapid calibration for 3-D freehand ultrasound. *Ultrasound in Medicine and Biology*, 24(6):855–869.

Pratt, P. and Arora, A. (2018). Transoral robotic surgery: image guidance and augmented reality. *ORL*, 80:204–212.

Pratt, P., Mayer, E., Vale, J., Cohen, D., Edwards, E., Darzi, A., and Yang, G.-Z. (2012). An effective visualisation and registration system for image-guided robotic partial nephrectomy. *Journal of Robotic Surgery*, 6(1):23–31.

Prokopetc, K., Collins, T., and Bartoli, A. (2015). Automatic detection of the uterus and fallopian tube junctions in laparoscopic images. In *International Conference on Information Processing in Medical Imaging*, pages 552–563.

Puerto-Souza, G. A., Cadeddu, J. A., and Mariottini, G. L. (2014). Toward long-term and accurate augmented-reality for monocular endoscopic videos. *IEEE Transactions on Biomedical Engineering*, 61(10):2609–2620.

Reaungamornrat, S., Liu, W. P., Wang, A. S., Otake, Y., Nithiananthan, S., Uneri, A., Schafer, S., Tryggestad, E., Richmon, J., Sorger, J. M., Siewerdsen, J. H., and Taylor, R. H. (2013). Deformable image registration for cone-beam CT guided transoral robotic base of tongue surgery. *Physics in Medicine & Biology*, 58(14):4951–4979.

Rosenthal, M., State, A., Lee, J., Hirota, G., Ackerman, J., Keller, K., Pisano, E. D., Jiroutek, M., Muller, K., and Fuchs, H. (2002). Augmented reality guidance for needle biopsies: an initial randomized, controlled trial in phantoms. *Medical Image Analysis*, 6(3):313–320.

Rousseau, F., Hellier, P., and Barillot, C. (2005). Confhusius: a robust and fully automatic calibration method for 3D freehand ultrasound. *Medical Image Analysis*, 9(1):25–38.

Rullier, E., Denost, Q., Vendrely, V., Rullier, A., and Laurent, C. (2013). Low rectal cancer: classification and standardization of surgery. *Diseases of the Colon and Rectum*, 56(5):560–567.

Schwartz, G. F., Goldberg, B. B., Rifkin, M. D., and D'Orazio, S. E. (1988). Ultrasonography: an alternative to x-ray-guided needle localization of nonpalpable breast masses. *Surgery*, 104(5):870–873.

Shen, J., Zemiti, N., Dillenseger, J.-l., and Poignet, P. (2018). Fast and simple automatic 3D ultrasound probe calibration based on 3D printed phantom and an untracked marker. pages 878–882.

Sigel, B., Spigos, D. G., Donahue, P. E., Pearl, R., Popky, G. L., and Nyhus, L. M. (1979). Intraoperative ultrasonic visualization of biliary calculi. *Current surgery*, 36(3):158–159.

Simpfendörfer, T., Baumhauer, M., Müller, M., Gutt, C. N., Meinzer, H.-P. P., Rassweiler, J. J., Guven, S., and Teber, D. (2011). Augmented reality visualization during laparoscopic radical prostatectomy. *Journal of Endourology*, 25(12):1841–1845.

Singla, R., Edgcumbe, P., Pratt, P., Nguan, C., and Rohling, R. (2017). Intra-operative ultrasound-based augmented reality guidance for laparoscopic surgery. *Healthcare Technology Letters*, 4(5):204–209.

Teber, D., Guven, S., Simpfendörfer, T., Baumhauer, M., Güven, E. O., Yencilek, F., Gözen, A. S., and Rassweiler, J. (2009). Augmented reality: a new tool to improve surgical accuracy during laparoscopic partial nephrectomy? preliminary in vitro and in vivo results. *European Urology*, 56(2):332–338.

Treeby, B. E. and Cox, B. T. (2010). k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields. *Journal of Biomedical Optics*, 15(2):021314.

Trotta, B. M., Pease, C. S., Rasamny, J. J., Raghavan, P., and Mukherjee, S. (2011). Oral cavity and oropharyngeal squamous cell cancer: key imaging findings for staging and treatment planning. *RadioGraphics*, 31(2):339–354.

Tsai, R. and Lenz, R. (1989). A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation*, 5(3):345–358.

Ukimura, O., Nakamoto, M., Sato, Y., Hashizume, M., Miki, T., Desai, M., Aron, M., and Gill, I. S. (2010). Augmented reality for image-guided surgery in urology. In *New Technologies in Urology*, pages 215–222.

Van Abel, K. M. and Moore, E. J. (2013). Surgical management of the base of tongue. *Operative Techniques in Otolaryngology - Head and Neck Surgery*, 24:74–85.

Vermaas, M., Ferenschild, F., Verhoef, C., Nuyttens, J., Marinelli, A., Wiggers, T., Kirkels, W., Eggermont, A., and de Wilt, J. (2007). Total pelvic exenteration for primary locally advanced and locally recurrent rectal cancer. *European Journal of Surgical Oncology*, 33(4):452–458.

Vitrani, M. A., Baumann, M., Reversat, D., Morel, G., Moreau-Gaudry, A., and Mozer, P. (2016). Prostate biopsies assisted by comanipulated probe-holder: first in man. *International Journal of Computer Assisted Radiology and Surgery*, 11(6):1153–1161.

Volonté, F., Pugin, F., Bucher, P., Sugimoto, M., Ratib, O., and Morel, P. (2011). Augmented reality and image overlay navigation with OsiriX in laparoscopic and robotic surgery: Not only a matter of fashion. *Journal of Hepato-Biliary-Pancreatic Sciences*, 18(4):506–509.

Wang, Q., Khanicheh, A., Leiner, D., Shafer, D., and Zobel, J. (2017). Endoscope field of view measurement. *Biomedical Optics Express*, 8(3):1441–1454.

Wengert, C., Reeff, M., Cattin, P. C., and Székely, G. (2006). Fully automatic endoscope calibration for intraoperative use. In *Bildverarbeitung für die Medizin*, pages 419–423, Berlin. Springer.

Wild, E., Teber, D., Schmid, D., Simpfendörfer, T., Müller, M., Baranski, A.-C., Kenngott, H., Kopka, K., and Maier-Hein, L. (2016). Robust augmented reality guidance with fluorescent markers in laparoscopic surgery. *International Journal of Computer Assisted Radiology and Surgery*, 11(6):899–907.

Yen, J. T., Steinberg, J. P., and Smith, S. W. (2000). Sparse 2-D array design for real time rectilinear volumetric imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 47(1):93–110.

Zhang, L., Ye, M., Chan, P.-L., and Yang, G.-Z. (2017). Real-time surgical tool tracking and pose estimation using a hybrid cylindrical marker. *International journal of computer assisted radiology and surgery*, 12(6):921–930.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22.

Appendix

The study in [Malti and Barreto, 2010] described the implementation of two popular hand-eye calibration methods [Tsai and Lenz, 1989; Daniilidis and Bayro-Corrochano, 1996] using quaternions or dual quaternion. The hand-eye calibration formulation $\mathbf{AX} = \mathbf{XB}$ can be written as:

$$\begin{pmatrix} \mathbf{R}_A & \overrightarrow{t_A} \\ \mathbf{0}_{1\times3} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R}_X & \overrightarrow{t_X} \\ \mathbf{0}_{1\times3} & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{R}_X & \overrightarrow{t_X} \\ \mathbf{0}_{1\times3} & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R}_B & \overrightarrow{t_B} \\ \mathbf{0}_{1\times3} & 1 \end{pmatrix}.$$

Tsai & Lenz's method

It finds the rotation $\mathbf{R}_X$ first, then, uses the rotation solution to compute the translation $\overrightarrow{t_X}$.

1. solving rotation $\mathbf{R}_X$:

   A quaternion $\mathbf{q}$ consists of a scalar q_0 and a 3D vector $\overrightarrow{q}$:

   $$\mathbf{q} = \begin{pmatrix} q_0 \\ \overrightarrow{q} \end{pmatrix}.$$

   $\mathbf{R}_A\mathbf{R}_X = \mathbf{R}_X\mathbf{R}_B$ can be represented by quaternion multiplication $\mathbf{a.q} = \mathbf{q.b}$ where $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{q}$ are the quaternions associated with $\mathbf{R}_A$, $\mathbf{R}_B$ and $\mathbf{R}_X$, respectively. Then, based on the definition of the product of two quaternions

$$\mathbf{a.b} = \begin{pmatrix} a_0 b_0 - \overrightarrow{a}^T . \overrightarrow{b} \\ a_0 \overrightarrow{b} + b_0 \overrightarrow{a} + \overrightarrow{a} \times \overrightarrow{b} \end{pmatrix},$$

the quaternion multiplication can be represented by

$$\underbrace{\begin{pmatrix} a_0 - b_0 & -(\overrightarrow{a} - \overrightarrow{b})^T \\ \overrightarrow{a} - \overrightarrow{b} & [\overrightarrow{a} + \overrightarrow{b}]_\times + (a_0 - b_0)\mathbf{I}_3 \end{pmatrix}}_{\mathbf{K(a,b)}} \mathbf{q} = 0.$$

The solution $\mathbf{q}$ is computed by the singular-value decomposition (SVD) of

$$L = \begin{pmatrix} \mathbf{K}(\mathbf{a}_1, \mathbf{b}_1) \\ \vdots \\ \mathbf{K}(\mathbf{a}_n, \mathbf{b}_n) \end{pmatrix}$$

and rotation $\mathbf{R}_X$ is the inverse conversion of the quaternion $\mathbf{q}$.

2. solving translation $\overrightarrow{t_X}$:
   Based on the solution of rotation $\mathbf{R}_X$, translation $\overrightarrow{t_X}$ is formulated as a non-constrained least square problem:

$$min \left\| \begin{pmatrix} \mathbf{R}_{A_1} - \mathbf{I}_3 \\ \vdots \\ \mathbf{R}_{A_n} - \mathbf{I}_3 \end{pmatrix} \overrightarrow{t_X} - \begin{pmatrix} \mathbf{R}_X \overrightarrow{t_{B_1}} - \overrightarrow{t_{A_1}} \\ \vdots \\ \mathbf{R}_X \overrightarrow{t_{B_n}} - \overrightarrow{t_{A_n}} \end{pmatrix} \right\|_2$$

Daniilidis' method

In this method, rotation and translation are computed at the same time. A homogeneous transformation can be described using a dual quaternion:

$$\widehat{\mathbf{q}} = \mathbf{q} + \varepsilon \mathbf{q}' = \begin{pmatrix} q_0 \\ \overrightarrow{q} \end{pmatrix} + \varepsilon \begin{pmatrix} q_0' \\ \overrightarrow{q}' \end{pmatrix}$$

where $\mathbf{q}$ and $\mathbf{q}$' are quaternions corresponding to rotation $\mathbf{R}$ and translation $\overrightarrow{t}$, respectively; $\varepsilon^2 = 0$ is a scalar constant.

1. Let $\widehat{\mathbf{a}}$, $\widehat{\mathbf{b}}$ and $\widehat{\mathbf{q}}$ be the unit dual quaternions associated with **A**, **B** and **X**, the hand-eye calibration can be represented by $\widehat{\mathbf{a}}.\widehat{\mathbf{q}} = \widehat{\mathbf{q}}.\widehat{\mathbf{b}}$
2. Split the dual quaternions multiplication in real and dual parts: $\mathbf{a}.\mathbf{q} = \mathbf{q}.\mathbf{b}$ and $\mathbf{a}'.\mathbf{q} + \mathbf{a}.\mathbf{q}' = \mathbf{q}.\mathbf{b}' + \mathbf{q}'.\mathbf{b}$
3. Based on the definition of the product of two quaternions

   $$\mathbf{a}.\mathbf{b} = \begin{pmatrix} a_0 b_0 - \overrightarrow{a}^T.\overrightarrow{b} \\ a_0 \overrightarrow{b} + b_0 \overrightarrow{a} + \overrightarrow{a} \times \overrightarrow{b} \end{pmatrix},$$

   the dual quaternions multiplication is represented by:

   $$\underbrace{\begin{pmatrix} \overrightarrow{a} - \overrightarrow{b} & |\overrightarrow{a} + \overrightarrow{b}|_\times & \mathbf{0}_{3\times 1} & \mathbf{0}_{3\times 3} \\ \overrightarrow{a}' - \overrightarrow{u_1}^T, \overrightarrow{u_1}^T b' & |\overrightarrow{a}' + \overrightarrow{b}'|_\times & \overrightarrow{a} - \overrightarrow{b} & |\overrightarrow{a} + \overrightarrow{b}|_\times \end{pmatrix}}_{\mathbf{M}} \begin{pmatrix} \mathbf{q} \\ \mathbf{q}' \end{pmatrix} = 0$$

4. Assuming $n \geqslant 2$ motions of endoscope, the global linear matrix is constructed

   as $\mathbf{P} = \begin{pmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_n \end{pmatrix}$, and decomposed by $\mathbf{P} = \mathbf{U}\Sigma\mathbf{V}^T$.

5. The right-singular vectors $\overrightarrow{ev_1}$ and $\overrightarrow{ev_2}$ corresponding to the minimum singular values are selected and represented by two 4×1 vectors

   $\overrightarrow{ev_1} = \begin{pmatrix} \overrightarrow{u_1} \\ \overrightarrow{v_1} \end{pmatrix}$ and $\overrightarrow{ev_2} = \begin{pmatrix} \overrightarrow{u_2} \\ \overrightarrow{v_2} \end{pmatrix}$. Since $\begin{pmatrix} \mathbf{q} \\ \mathbf{q}' \end{pmatrix}$ is a linear combination of the right singular vectors, $\begin{pmatrix} \mathbf{q} \\ \mathbf{q}' \end{pmatrix} = \lambda_1 \begin{pmatrix} \overrightarrow{u_1} \\ \overrightarrow{v_1} \end{pmatrix} + \lambda_2 \begin{pmatrix} \overrightarrow{u_2} \\ \overrightarrow{v_2} \end{pmatrix}$.

6. The constraints $\mathbf{q}^T\mathbf{q} = 1$ and $\mathbf{q}^T\mathbf{q}' = 0$ imply two quadratic equations to determine λ_1 and λ_2, thereby finding the solutions of **q** and **q**'.
7. The inverse conversions of **q** and **q**' are respectively corresponding to rotation R_X and translation vector $\overrightarrow{t_X}$.

Malti and Barreto's method

This method deduces the quaternion $\mathbf{q}$ by computing the rotation R_X and translation $\overrightarrow{t_X}$ respectively using Tsai and Lenz's method and the dual part of the dual quaternions multiplication in Daniilidis's method:
$\mathbf{K}(\mathbf{a}', \mathbf{b}')\mathbf{q} + \mathbf{K}(\mathbf{a}, \mathbf{b})\mathbf{q}' = 0$. Assuming $n \geqslant 2$ motions of endoscope, then $L'\,\mathbf{q} + L\,\mathbf{q}' = 0$ with

$$L = \begin{pmatrix} \mathbf{K}(\mathbf{a}_1, \mathbf{b}_1) \\ \vdots \\ \mathbf{K}(\mathbf{a}_n, \mathbf{b}_n) \end{pmatrix} \text{ and } L' = \begin{pmatrix} \mathbf{K}(\mathbf{a}'_1, \mathbf{b}'_1) \\ \vdots \\ \mathbf{K}(\mathbf{a}'_n, \mathbf{b}'_n) \end{pmatrix}.$$

By solving the least squares probem of $min\ \|L'\,\mathbf{q} + L\,\mathbf{q}'\|_2$ obeying a constraint of $q_0\ q'_0 + <\overrightarrow{q}, \overrightarrow{q}'>= 0$, the dual component $\mathbf{q}'$ of the dual quaternion can be computed. It is used to calculate the translation $\overrightarrow{t_X}$ based on following relation $\begin{pmatrix} 0 \\ \overrightarrow{t_X} \end{pmatrix} = 2\,\mathbf{q}'.\ \mathbf{q}^*$ with conjugate of $\mathbf{q}$.

List of publications

Shen et al., 2019a Shen J., Zemiti N., Dillenseger J.-L., Poignet P. (2019). Intraoperative ultrasound-based augmented reality guidance. *Surgetica 2019*.

Shen et al., 2019b Shen J., Zemiti N., Taoum C., Aiche G., Dillenseger J.-L., Rouanet P., Poignet P. (2019). Transrectal ultrasound image-based real-time augmented reality guidance in robot-assisted laparoscopic rectal surgery: a proof-of-concept study. *International Journal of Computer Assisted Radiology and Surgery (IJCARS)* doi:10.1007 /s11548-019-02100-2.

Shen et al., 2018a Shen J., Zemiti N., Viquesne, A., Caravaca-Mora, O., Courtin A., Garrel R., Dillenseger J.-L., Poignet P. (2018). Intraoperative ultrasonography-based augmented reality for application in image guided robotic surgery. Computer *32nd Assisted Radiology and Surgery (CARS), International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, 13, Supplement 1: S45-S46.

Shen et al., 2018b Shen J., Zemiti N., Dillenseger J.-L., Poignet P. (2018). Fast and simple automatic 3D ultrasound probe calibration based on 3D printed phantom and an untracked marker. *40th IEEE Conference on Engineering in Medicine and Biology Conference (EMBC)*, pages 878-882.

Shen et al., 2017a Shen J., Zemiti N., Caravaca, O., Simon A., Dillenseger J.-L., Poignet P. (2017). Intraoperative 3D ultrasound probe calibration. *Recherche en Imagerie et Technologies pour la Santée (RITS) 2017*.

Shen et al., 2017b Shen J., Courtin A., Viquesne, A., Zemiti N., Caravaca, O., Simon A., Dillenseger J.-L., Poignet P. (2017). Augmented perception in transoral robotic surgery for tongue base cancer. *7th Joint Workshop on New Technologies for Computer/Robot Assisted Surgery (CRAS) 2017.*

Shen et al., 2017c Shen J., Zemiti N., Caravaca, O., Simon A., Dillenseger J.-L., Poignet P. (2017). Augmented reality visualization based on 3D ultrasonography. *Surgetica 2017.*

UNIVERSITE
BRETAGNE
LOIRE / MATHSTIC

UNIVERSITÉ DE
RENNES 1

Titre : Mise en place d’une solution de réalité augmentée pour le guidage du geste en chirurgie robotisée à partir d’échographies : application à la chirurgie transorale et la chirurgie laparoscopicque gastro-intestinale

Mot clés : Réalité augmentée, échographie, calibration de sondes d’échographie, guide de résection tumorale, chirurgie robotique

Résumé : Cette thèse porte sur le développement d’une solution de réalité augmentée dans le cadre de la chirurgie robotisée et plus particulièrement pour la chirurgie transorale des tumeurs de la base de langue et la chirurgie laparoscopique des cancers du bas-rectum. Une des problématiques pour les chirurgiens est de repérer sur la vue endoscopique les limites de la tumeur et les marges de résections. Celles-ci sont en effet non visibles directement. L’échographie peropératoire est largement utilisée pour repérer les tumeurs lors des interventions. Nous proposons donc une solution de réalité augmentée dans laquelle l’information extraite de l’échographie est reprojetée sur la vision binoculaire de la station de chirurgie robotisée afin de guider le chirurgien dans la résection de la tumeur. Plusieurs verrous de cette chaîne de traitement ont été repérés et étudiés. Nous avons ainsi proposé une nouvelle méthode pour la calibration de sondes d’échographie. Nous avons démontré que cette méthode était plus facile à mettre en œuvre, plus rapide et plus précise que les méthodes proposées dans la littérature. Cette sonde calibrée, associée à des outils de localisation et de calibration de la sonde endoscopique nous a permis de proposer une solution de réalité augmentée qui permettait de reprojeter l’information acquise sur l’image sur la vue endoscopique avec des erreurs inférieures à 1 mm. Nous avons alors établi la preuve de concept de l’application de cette chaîne de réalité augmentée dans deux expérimentations, l’une sur un fantôme physique en silicone du rectum et l’autre sur une langue de mouton en ex-vivo. Les résultats expérimentaux ont montré que l’information augmentée avait permis au chirurgien de percevoir avec précision les marges de résections des tumeurs simulées et d’accomplir le geste opératoire à l’aide de cette perception.

Title: Framework for Ultrasonography-based Augmented Reality in Robotic Surgery: Application to Transoral Surgery and Gastrointestinal Surgery

Keywords: Augmented reality, ultrasonography, ultrasound probe calibration, tumor resection guidance, robotic surgery

Abstract: The medical context of this thesis is transoral robotic surgery for base of tongue cancer and robot-assisted laparoscopic surgery for low-rectal cancer. One of the main challenges for surgeons to perform these two surgical procedures is to identify the tumor resection margins accurately, because tumors are often concealed in base of tongues or rectal walls and there is lack of efficient intraoperative guidance systems. However, ultrasonography is widely used to image soft-tissue tumors, which motivates our proposition of an augmented reality framework based on intraoperative ultrasonography images for tumor resection guidance. The framework, proposed, with clinical partners, consists to adapt to the surgical workflow of robot-assisted surgery for treating base of tongue cancer and low-rectal cancer. For this purpose, we developed a fast and accurate 3D ultrasound probe calibration method to track the probe and facilitate its intraoperative use. Moreover, we evaluated the performance of the proposed framework augmenting an intraoperative endoscopic camera with ultrasound information, which shows less than 1mm error. Furthermore, we designed experimental protocols using a silicone rectum phantom and an ex-vivo lamb tongue, that simulate the integration of the implemented framework into the current surgical workflow. The experimental results show that, according to the augmented endoscopic views provided by the proposed framework, a surgeon is able to accurately identify the resection margins of the simulated tumors in these phantoms.