



**HAL**  
open science

# Image-based guidance of a transesophageal HIFU probe for the treatment of cardiac arrhythmias

Batoul Dahman

► **To cite this version:**

Batoul Dahman. Image-based guidance of a transesophageal HIFU probe for the treatment of cardiac arrhythmias. Signal and Image processing. Rennes 1, 2022. English. NNT : . tel-03892658

**HAL Id: tel-03892658**

**<https://theses.hal.science/tel-03892658>**

Submitted on 9 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1

ECOLE DOCTORALE N° 601

*Mathématiques et Sciences et Technologies  
de l'Information et de la Communication*

Spécialité : Signal, Image, Vision

Par

**Batoul Dahman**

## **Image-based guidance of a transesophageal HIFU probe for the treatment of cardiac arrhythmias**

Thèse présentée et soutenue à Rennes, le 6 octobre 2022

Unité de recherche : Laboratoire Traitement du Signal et de l'Image (LTSI), UMR Inserm 1099

### **Rapporteurs avant soutenance :**

María J. Ledesma Carbavo  
Denis Koumè

Professeur d'Université, BIT, Université Polytechnique de Madrid  
Professeur d'Université, IRIT, Université Paul Sabatier, Toulouse

### **Composition du Jury :**

Présidente : Diana Mateus

Examineurs : María J. Ledesma Carbavo  
Denis Koumè  
Diana Mateus  
Cyril Lafon  
Mireille Garreau

Professeur d'Université, BIT, Université Polytechnique de Madrid  
Professeur d'Université, IRIT, Université Paul Sabatier, Toulouse  
Professeur d'Université, LS2N, École Centrale Nantes, Nantes  
Directeur de Recherche, Inserm, LabTAU, Inserm Lyon  
Professeure d'Université, LTSI, Université de Rennes 1

Dir. de thèse : Jean-Louis Dillenseger

Maître de Conférences (HdR), LTSI, Université de Rennes 1







# Remerciement

Je voudrais tout d'abord remercier mon directeur de thèse Jean-Louis Dillenseger, qui m'a motivé à planifier et à développer mon projet de recherche. Ses remarques et corrections opportunes ont encadré ma thèse en la conduisant à une conclusion heureuse et valorisante.

Je remercie María J. Ledesma Carbavo et Denis Koumè qui ont acceptés d'être rapporteurs de ce manuscrit. Leurs commentaires ont amélioré la qualité de ce travail. De même, les discussions et les idées exprimés lors de la soutenance par les autres membres du jury, Diana Mateus, Cyril Lafon et Mireille Garreau qui me donnent le ton pour les futures publications et perspectives.

Je pense également à mes amis et mes collègues du laboratoire et les remercie pour leur aide, leur soutien scientifique et technique, et surtout leur bonne humeur et leur amabilité.

Toute ma gratitude et mes remerciements sont pour ceux qui ont toujours été présents malgré la distance, ceux qui m'ont fait confiance, ceux qui me donnait l'espoir, l'encouragement et l'enthousiasme. J'exprime ma gratitude à mes biens aimés parents, mon cher frère Hasan, ma grand-mère Nada, mon grand-père Ali et tous mes chers proches.

Je remercie du fond du cœur mon amour Alaa, mon cher mari, qui remplit ma vie d'amour, d'espoir et de bonheur, et qui m'a soutenu et m'a supporté tout au long de ma thèse. Merci et je t'aime.

Finalement, je remercie ma petite princesse Julia, qui a rempli notre vie d'amour, de joie et de bonheur. Je t'aime plus que ma vie.



# Résumé étendu de la Thèse en français

## Contexte de l'étude

Les travaux de cette Thèse portent sur le guidage par l'image d'un traitement de la fibrillation cardiaque, et plus particulièrement la fibrillation ventriculaire, par Ultrasons Focalisés Haute Intensité (HIFU) par voie trans-œsophagienne.

En deux mots, la fibrillation ventriculaire est une arythmie du cœur provenant d'un dysfonctionnement de la voie de conduction électrique dans le tissu cardiaque. La fibrillation ventriculaire se produit lorsque les signaux électriques qui indiquent au muscle cardiaque de pomper imposent aux ventricules de se contracter à très haute fréquence et de façon désordonnée (fibrillation). La fibrillation fait en sorte que le sang n'est pas pompé vers le corps. De plus, la fibrillation ventriculaire prolongée peut entraîner un arrêt cardiaque et la mort.

Pour les patients souffrant d'arythmies chroniques, des médicaments destinés à régulariser le rythme cardiaque ou pour rétablir un rythme normal sont généralement prescrits. Éventuellement, un stimulateur ou un défibrillateur automatique sont posés chirurgicalement. En cas de fibrillations réfractaires aux traitements, une ablation cathétérisée peut être envisagée. Actuellement, c'est l'ablation par radiofréquence (RF) qui est le traitement de référence. L'ablation consiste à éliminer par nécrose les tissus responsables des arythmies. Pour cela, un cathéter est inséré par voie fémorale ou sous clavière et est monté sous guidage fluoroscopique vers la cible. L'antenne RF est mise en contact de la cible et l'émission des ondes radiofréquences échauffent et nécrosent les tissus cardiaques sous-jacents. Cette thérapie a plusieurs inconvénients : 1) le geste est tout de même assez invasif, 2) le guidage vers la paroi cible sous fluoroscopie est assez compliquée, 3) du fait des mouvements cardiaques la pointe l'émetteur peut ne pas être en contact avec la paroi et donc une lésion transmurale n'est pas assurée entraînant un échec de la thérapie, 4) le fait de délivrer de l'énergie en direction de l'extérieur du cœur peut entraîner des lésions graves sur les organes environnants (œsophage, ...). Une technique d'ablation alternative par Ultrasons Haute Intensité Focalisés (HIF) par voie trans-oesophagienne a été proposée pour compenser les limitations évoquées précédemment. En effet, chez les humains, l'œsophage est placé juste derrière le cœur et offre donc une très bonne fenêtre acoustique pour les ultrasons (l'échographie cardiaque trans-oesophagienne exploite cette fenêtre). De surcroît, le traitement vers l'intérieur du cœur ne présente aucun risque pour les organes environnants.





Figure 1 –Traitement par HIFU de la fibrillation ventriculaire par voie trans-oesophagienne.

Un premier projet ANR (ANR CardioUSgHIFU) a permis de réaliser et de valider une première sonde trans-oesophagienne pour le traitement des fibrillations auriculaires (Figure 1) [1].

La sonde de thérapie comportait en son milieu une sonde d'imagerie qui fournissait une image échographique perpendiculaire à l'axe de l'œsophage. Une première solution de guidage par l'image a d'ailleurs été proposée lors de ce projet [Thèse Sandoval et PMB18]. Ce premier projet a permis de réaliser une preuve de concept de cette thérapie potentielle.

Les travaux développés lors de ce travail de ma Thèse ont été effectués dans le cadre du projet ANR CHORUS (ANR 17-CE19-0017) qui faisait suite au projet cardioUSgHIFU. Les objectifs de ce projet ont été justement de proposer l'instrumentation et de réaliser une validation préliminaire des approches d'ablation par HIFU pour le traitement de la fibrillation ventriculaire par voie trans-oesophagienne.

Le projet était décomposé en différentes parties :

- 1) Le développement d'une nouvelle sonde dual-mode permettant a) l'ablation de la paroi ventriculaire par focalisation géométrique et électronique et d'imager l'anatomie sur deux coupes perpendiculaires (Figure 2).

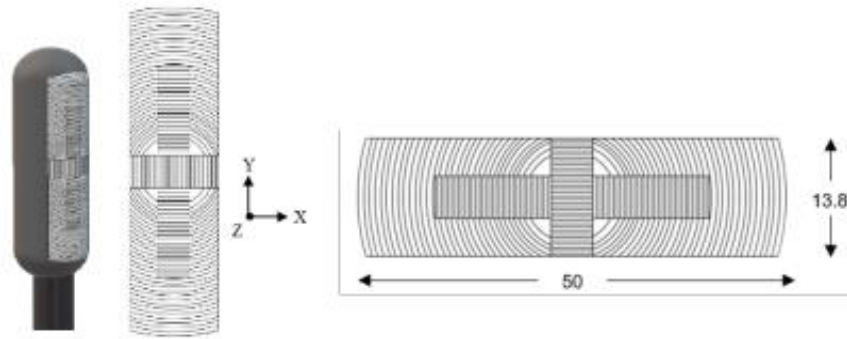


Figure 2 – Représentation schématique de la sonde avec 32 anneaux de thérapie deux barrettes d'éléments dual mode permettant d'acquérir deux plans de vue perpendiculaires en échographie.

- 2) La validation de la thérapie sur modèle animal ou modèle cœur isolé battant (Langendorff).
- 3) La faisabilité du guidage de la thérapie par l'image. Mon travail de Thèse a porté sur ce dernier point.

## Objectif de l'étude et organisation du travail de Thèse

Comme évoqué précédemment, l'objectif de cette thèse est de proposer des techniques de mise en correspondance entre l'imagerie préopératoire (volume scanner X) utilisée pour établir la planification de l'intervention et l'imagerie per-opératoire fournie par la sonde (images 2D échographique perpendiculaire à l'axe de la sonde). Ceci en utilisant la seule information fournie par les images sans solutions de tracking extérieur. Dans ce cas, le recalage consiste à estimer la pose 3D (position et orientation) de l'image ultrasonore (donc de la sonde) dans le volume CT 3D préopératoire.

L'étude menée lors d'une thèse précédente [Sandoval] a été basée sur certaines hypothèses très fortes : 1) Malgré le fait que le cœur soit un organe mobile une hypothèse de recalage rigide pouvait être retenue, car nous avons un ciné-scan (20 volumes acquis dans 20 phases du cycle cardiaque) et les images échographiques étaient synchronisées sur l'ECG. Ceci permet d'associer dans une même phase l'image échographique et le volume correspondant. De plus, vues de l'œsophage, les mouvements respiratoires subis par le cœur sont très faibles. Un recalage rigide pouvait être envisagé. 2) L'œsophage a une position contrainte par les organes et tissus qui l'entourent, tels que la colonne vertébrale, la trachée, la vascularisation carotido-jugulaire, l'arc aortique, l'artère pulmonaire droite, la bronche principale gauche, l'oreillette gauche et le diaphragme. L'hypothèse est alors que c'est l'œsophage qui contraint la trajectoire de la sonde HIFU et donc, dans le cas où la sonde d'imagerie produit une image perpendiculaire à l'axe de la sonde, que les images échographiques 2D sont perpendiculaires à l'axe de l'œsophage.

Ces deux hypothèses ont permis de proposer une solution de mise en correspondance (recalage) de l'échographie 2D et du Scanner X 3D basée sur (Figure 3) : 1) la segmentation de l'œsophage sur les données Scanner X préopératoires, 2) l'extraction des coupes scanner X perpendiculaires à l'axe de l'œsophage (de même orientation que l'échographie 2D) ; 3) le recalage 2D/2D entre échographie 2D et coupes 2D extraites du scanner X et, 4) le choix du couple échographie/coupe

scanner le plus ressemblant permettait d'estimer le recalage 2D/3D final et donc de reporter l'information contenue dans le scanner X (cible) sur l'échographie 2D.

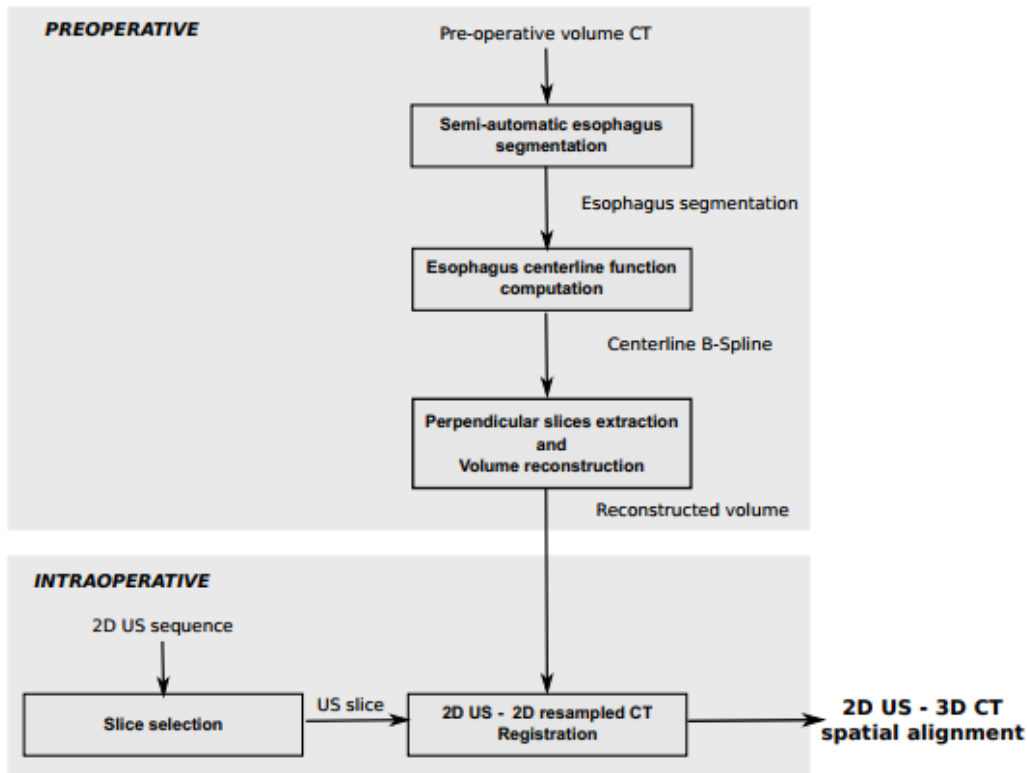


Figure 3 – Principe du recalage échographie 2D/scanner X 3D développé dans le cadre du projet CardioUSgHIFU [2].

Cette méthodologie a permis l'estimation de la pose de la sonde dans le volume scanner avec erreurs médianes de recalage de cible (Target Registration Error) de l'ordre de 5 mm.

Mon travail de Thèse consiste alors à intégrer une nouvelle configuration d'imageries liée au nouveau projet, à relâcher certaines contraintes liées aux hypothèses un peu trop fortes et à accélérer le temps de calcul du recalage pour pouvoir envisager une utilisation en routine clinique. Plus particulièrement :

- Nous avons amélioré la solution itérative précédente sur deux aspects (chapitre 2 de la thèse) :
  - J'ai intégré la nouvelle configuration d'imagerie, à savoir 2 plans images perpendiculaires (cf. Figure 2) dans la solution de recalage développée précédemment (chapitre 2 de la thèse).
  - En considérant que la technique classique précédente donne une première estimée de la pose, nous avons relâché la contrainte anatomique qui imposait une image échographique perpendiculaire à l'axe de l'œsophage en recherchant la pose de l'image dans l'environnement 3D autour de la première estimée. En d'autres mots nous avons fait un recalage spatial direct entre la paire d'échographies 2D/ et le scanner 3D sans passé par des coupes intermédiaires.

- La solution précédente de recalage itératif est basée sur une série de recalage échographies 2D/ coupes scanner 2D (étape 3 de la méthodologie décrite précédemment). Ce recalage était effectué en utilisant une méthode classique itérative qui prenait 6 s par paire d'images. L'idée est alors d'utiliser des méthodes de recalage par apprentissage profond qui ont la particularité d'accélérer grandement les temps de calcul tout en préservant (ou en améliorant) la précision du recalage. Nous avons donc proposé une méthode de recalage rigide échographie 2D/coupe scanner X 2D par apprentissage profond avec apprentissage supervisé (chapitre 3 de la thèse). Ce modèle a été ensuite étendu pour réaliser un recalage rigide échographie 2D/volume scanner X 3D par apprentissage profond avec apprentissage supervisé (chapitre 3 de la thèse).
- Nous avons voulu ensuite relâcher la contrainte de recalage rigide. Si nous supposons que nous avons une très bonne estimée de la pose de sonde, un recalage élastique peut alors être considéré. Nous avons donc proposé méthode de recalage élastique entre échographie 2D/coupe scanner X 2D par apprentissage profond avec apprentissage non-supervisé.

Avant de présenter ces méthodes je voudrai faire un point sur l'évaluation et, le cas échéant, l'apprentissage des méthodes. En imagerie biomédicale, il est extrêmement difficile d'obtenir des vérités terrains sur les transformations subies entre images à recaler, particulièrement si les modalités sont différentes et/ou si les transformations ne sont pas rigides (d'où la popularité des méthodes avec apprentissage non supervisé). Dans notre cas nous avons un second souci, car la sonde dual mode avec deux plans images échographiques perpendiculaires n'était pas encore développée au moment de la Thèse. Nous avons donc décidé d'utiliser un simulateur qui pouvait générer des images échographiques à partir de données scanner X, des caractéristiques acoustiques des tissus (impédance acoustique et distributions spatiales de réflecteurs générant le speckle) et des caractéristiques de la sonde (nombre d'éléments, fréquence d'émission, bande passante, forme du champ ultrasonore) [3]. Ce simulateur permettait de créer des vérités terrains utilisés soit pour l'apprentissage des méthodes de deep learning, soit pour l'évaluation de ces méthodes. Ce simulateur a été appliqué sur de nombreux volumes scanner X et avec différents paramètres de simulation.

### **Estimation de la pose de deux plans échographiques perpendiculaires dans le scanner X préopératoire par une technique itérative.**

La nouvelle sonde qui va développer dans le cadre du projet Chorus permet d'obtenir deux plans perpendiculaires d'images échographiques (Figure 2). Le guidage de la thérapie nécessite d'estimer la pose de ces images échographiques dans le scanner X préopératoire. À partir d'une première estimée de la pose de la sonde obtenue par exemple en utilisant les travaux de Sandoval [2], l'idée est d'une part de faire une recherche de la transformée 3D autours de cette pose (cela permet de relâcher la contrainte anatomique de perpendicularités par rapport à l'œsophage) et d'apporter une

information spatiale supplémentaire en ajoutant un second plan d'image échographique perpendiculaire au premier.

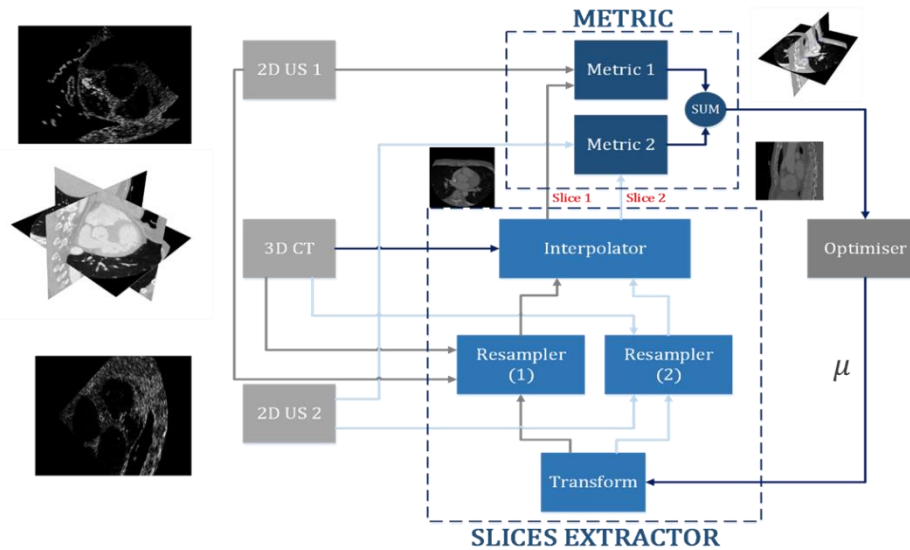


Figure 4 – Schéma de principe du recalage de 2 plans échographiques perpendiculaires et un volume scanner X 3D.

Le schéma de principe du recalage entre 2 plans échographiques perpendiculaires et un volume scanner X reprend le processus itératif classique : 1) à partir d'une pose estimée, on extrait deux plans de coupes perpendiculaires dans le volume scanné X, 2) une mesure de similarité est appliquée pour comparer la ressemblance entre les deux coupes scanner X extraites et les deux images échographiques et 3) un optimiseur essaye de modifier itérativement les paramètres de la pose afin de maximiser la ressemblance entre coupes scanner et images échographiques.

Dans notre cas précis, la transformation géométrique entre images est une transformation rigide, la mesure de similarité choisie est l'Information Mutuelle (cette méthode a été choisie à la suite d'une étude prospective précédente [4]) et l'optimiseur choisi est la descente de gradient stochastique.

Une évaluation a été menée pour estimer l'apport de la coupe perpendiculaire dans la précision de l'estimation de la pose. Les images utilisées par notre évaluation ont été d'une part un volume scanner X clinique et d'autre part des paires d'images échographiques obtenues par simulations et dont les poses dans le scanner X étaient connues. Concernant l'estimation des paramètres de la transformation le fait d'utiliser deux coupes perpendiculaires au lieu d'une permettait de diminuer l'erreur médiane de l'estimée de la translation 3D de 1,5 mm à 0,7 mm et l'estimée de l'angle de rotation 3D de 3° à 2,1°. Nous avons également mesuré des erreurs en distance de recalage (Target Registration Errors) sur certains points fiduciels 3D connus. Là encore l'erreur médiane a diminué de 2,54 à 1,7 mm en ajoutant un second plan image échographique.

## Méthode de recalage rigide échographie 2D/coupe scanner X 2D par apprentissage profond avec apprentissage supervisé.

L'idée de ce travail est de remplacer la procédure itérative de recalage rigide entre échographies 2D et coupes 2D extraites du scanner X (Figure 3) par une procédure plus rapide de recalage basé sur de l'apprentissage profond. Les résultats attendus sont une accélération du temps de calcul pour la rendre utilisable en routine clinique tout en préservant ou améliorant la précision du recalage. Comme énoncé précédemment, nous sommes capables de générer par simulation des images échographiques à partir des coupes scanner X. Nous pouvons donc générer des paires d'images avec des transformations connues. Un apprentissage supervisé peut alors être envisagé. Comme l'information entre les deux modalités est très différente (différence d'impédances acoustiques et speckle pour l'échographie et coefficient d'atténuations aux rayons X pour le scanner) nous avons envisagé une procédure de recalage avec différents actions (Figure 5) :

- 1) Un réseau siamois est appliqué sur les images à recaler. Ce réseau est composé deux sous-réseaux identiques, appelés réseaux jumeaux, d'architecture et de poids identiques Ils travaillent en parallèle et sont chargés de créer des représentations vectorielles pour les entrées. Ils aident à produire de meilleures représentations vectorielles en mesurant les similitudes entre les vecteurs. En sortie, nous avons deux cartes de caractéristiques (une par image d'entrée), et qui sont analogues à des descripteurs locaux denses. Dans notre cas, les réseaux jumeaux sont basés sur le modèle ResNet18 réputé pour sa bonne performance en extraction de caractéristiques d'images. Nous avons utilisé le modèle ResNet pré-entraîné sur ImageNet.
- 2) La concaténation des deux cartes de caractéristiques pour pouvoir servir d'entrée à :
- 3) Un réseau consultatif de recalage qui estime directement l'ensemble des paramètres de la transformation rigide (deux translations, une rotation). Dans notre cas, notre réseau est composé de trois blocs de couches convolutionnelles utilisant un noyau de taille 5, chacune suivie de couches de normalisation par lots, et d'une unité linéaire rectifiée (ReLU). La dernière couche est une couche entièrement connectée permettant d'estimer les paramètres de recalage rigide.

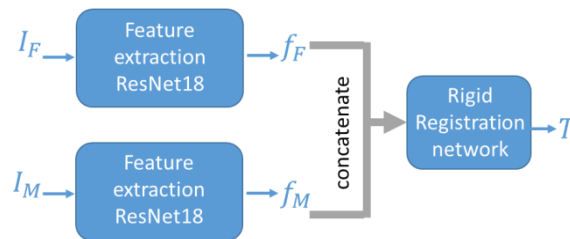


Figure 5 – Procédure de recalage rigide échographie 2D ( $I_F$ ) coupes scanner X ( $I_M$ ). Un réseau siamois extrait des caractéristiques des images et un réseau de recalage estime directement les paramètres géométriques de la transformation rigide ( $T$ ).

Le fait de connaître les images et la transformation géométrique entre-elles, nous a permis un apprentissage de bout en bout. La fonction de perte est la norme L2 entre les paramètres de recalage estimés et les paramètres recalage de la vérité terrain.

Les données utilisées pour l'apprentissage et l'évaluation sont issus de 20 volumes de scanner X cardiaques [5], [6]. 18 volumes sont utilisés pour l'apprentissage et 2 pour l'évaluation. Pour chaque volume, nous choisissons au hasard 200 positions initiales le long des axes de l'œsophage et extrayons les coupes obliques perpendiculaires à l'œsophage. Pour chaque position initiale nous plaçons aléatoirement la pose de notre échographe dans une plage de 10 mm en translation et 15 degrés en rotation et nous simulons l'image échographique. Nous avons donc 3600 paires d'images avec vérité terrain pour l'apprentissage et 400 pour l'évaluation.

Sur les 400 paires, nous avons comparé les performances de notre méthode par rapport à l'algorithme itératif classique (celui implémenté dans SimpleITK en utilisant l'information mutuelle en mesure de similarité) en termes de précision de recalage (erreur d'estimation des paramètres et erreurs de recalage de points fiduciaires -TRE) et en temps de calcul. Les médianes des erreurs d'estimation des paramètres étaient du même ordre de grandeur (voire un peu meilleures mais statistiquement non démontrées) avec notre méthode comparée à l'algorithme itératif (1,1 mm vs. 1,2 mm pour la translation et 2,1° vs. 2,4°). Cette tendance est vérifiée sur les TREs mesurés sur 8 points fiduciaires par image (médianes des TREs de 2,2 mm vs. 2,7 mm). Par contre, le gain en temps de calcul répond bien à nos attentes : 3 ms par paire d'images comparé à 6 s pour la méthode itérative.

Dans la section suivante, nous intégrerons l'approche d'apprentissage des caractéristiques à une procédure HIFU non invasive pour améliorer la planification et l'orientation de la thérapie. Nous appliquerons notre approche sur un recalage basé sur l'apprentissage 2D/3D pour affiner l'estimation du placement de la pose de la sonde trans-œsophagienne dans le volume préopératoire 3D.

### **Méthode de recalage rigide échographie 2D/Volume scanner X 3D par apprentissage profond avec apprentissage supervisé.**

La méthode précédente a permis de montrer que le recalage itératif classique 2D/2D pouvait être remplacé par une méthode par apprentissage profond, même avec des images de natures extrêmement différentes comme le scanner X et l'échographie. Cette méthode de recalage peut alors être intégrée dans le schéma d'estimation de la pose 3D de la sonde échographique proposé par précédemment dans notre laboratoire par Sandoval [7]. Par contre une des limites de la méthode de Sandoval est la contrainte forte que les images échographiques doivent être strictement perpendiculaires à l'axe de l'œsophage. Afin de relâcher cette contrainte, et suite à notre montée en compétences dans le domaine de l'apprentissage profond, nous avons élaboré une première preuve de concept pour résoudre directement le problème complexe de l'estimation de la pose 3D d'une coupe échographique dans le volume (ou un sous-volume) 3D.

Paramètres de la transformation rigide (3 translations, 3 rotations) de la pose de l'image échographique dans le sous-volume scanner X.

Comme pour le cas 2D/2D, nous avons décidé de séparer le schéma global en deux sous-problèmes (Figure 6).

- 1) L'extraction des caractéristiques des données d'entrée. Le principal défi était de concevoir les réseaux parallèles d'extraction de caractéristiques pour des données de dimensions différentes. Comme le recalage est 3D, les caractéristiques doivent être décrites dans un volume 3D. Nous avons conçu un réseau pour traiter la coupe échographique d'abord en 2D suivi d'une extension vers des couches 3D. Le sous-volume scanner est lui traité par un réseau 3D directement. Les données 3D issues de ces deux réseaux parallèles sont ensuite concaténées.
- 2) Un réseau de neurones qui va estimer directement les 6 paramètres de la transformation) partir des caractéristiques 3D extraites lors de l'étape précédente. Comme pour le cas 2D/2D, nous avons utilisé un modèle Resnet comme réseau de recalage pour estimer les paramètres.

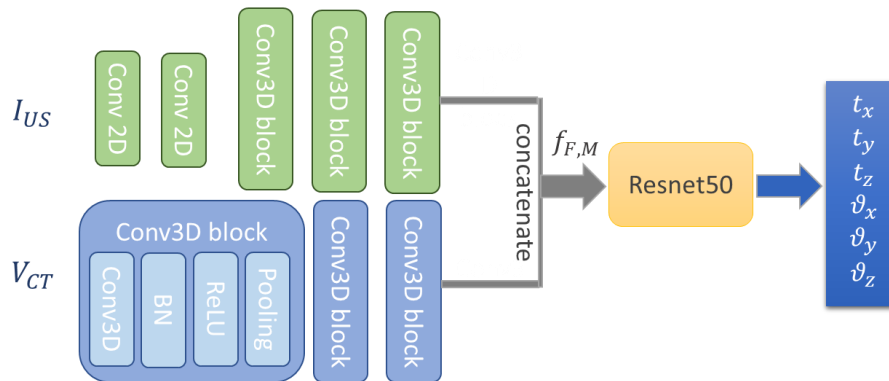


Figure 6 – Procédure de recalage rigide échographie ( $I_{US}$ ) / volume scanner X ( $V_M$ ). Un réseau parallèle extrait des caractéristiques 3D des images/volumes d'entrée et un réseau de recalage estime directement les paramètres géométriques de la transformation rigide (T).

Dans les conditions réelles de la thérapie, nous avons une idée approximative de la position de la pointe de l'endoscope (longueur insérée, analyse visuelle de la séquence d'images pendant la navigation, fluoroscopie, etc.). Ceci nous permet de définir une zone candidate le long de l'axe de l'œsophage dans laquelle le capteur d'images échographiques peut être situé. La taille de cette zone est d'environ 10 mm le long de l'œsophage. Nous sommes ainsi capables d'extraire un sous-volume de taille  $512 \times 512 \times 32$  voxels dans lequel se trouvera l'image US 2D. Ce sous-volume servira de volume d'entrée à notre réseau. Le fait d'estimer les paramètres dans un sous-volume a plusieurs avantages : il permet de réduire l'espace de recherche du réseau pour trouver la transformation optimale. Il réduit également la charge de la mémoire pendant la phase de formation et une stratégie d'augmentation des données.

Comme nous ne disposons pas de vérité terrain, nous avons décidé d'apprendre et d'évaluer notre réseau sur des volumes CT 3D réels et des images US 2D simulées.

Les volumes scanner sont issus de la base de données MMWHS2017 [5], [6], qui contient 60 volumes CT 3D. 70 % de l'ensemble de données est sélectionné au hasard de manière aléatoire



pour l'entraînement et les 30 % restants sont utilisés pour les tests. Pour chaque volume, nous avons extrait de manière aléatoire 3 sous-volumes de taille  $512 \times 512 \times 32$  voxels. Pour chaque sous-volume nous avons défini une position centrale (le centre de l'œsophage dans la coupe centrale du sous-volume) autour de laquelle nous avons défini de manière aléatoire la pose de l'image échographique en appliquant une transformation aléatoire dans une plage de  $\pm 10$  mm en translation et  $\pm 5^\circ$  en rotation autour de chaque axe de coordonnées. Le plan de coupe oblique défini par cette pose dans le sous-volume scanner X servira alors pour la simulation de l'image échographique. Au final, nous disposons de 126 paires d'images US/sous-volumes CT pour l'apprentissage et 54 pour le test.

Très peu d'articles ont utilisé une approche basée sur l'apprentissage pour le recalage de la coupe au volume, et plus précisément, nous n'avons trouvé aucun travail effectué récemment pour notre application spécifique. Nous avons donc décidé de comparer nos résultats à la méthode itérative classique (avec MI comme métrique de similarité) utilisant la bibliothèque simpleITK. Nous avons comparé les résultats de notre méthode à la méthode itérative classique en termes de précision de recalage (la distance moyenne entre le plan estimé et le plan réel -DistErr-, et les erreurs sur chaque paramètre) et de temps de calcul (Nous constatons bien une accélération du temps de calcul 0,07 seconde pour notre méthode (presque 140 fois moins que la méthode classique). Cette accélération n'a pas été obtenue au détriment de la précision du recalage car avec notre méthode, cette précision est du même ordre de grandeur voire légèrement meilleure que pour la méthode classique.

Table 1)

Nous constatons bien une accélération du temps de calcul 0,07 seconde pour notre méthode (presque 140 fois moins que la méthode classique). Cette accélération n'a pas été obtenue au détriment de la précision du recalage car avec notre méthode, cette précision est du même ordre de grandeur voire légèrement meilleure que pour la méthode classique.

Table 1: Performances de notre méthode (CNN) comparées à la méthode classique (SimpleITK).

Méthode	DisErr (mm)	Erreur moyenne de l'estimation des paramètres de transformation en mm pour les translations et en $^\circ$ pour les rotations.						Temps (Sec)	
			$t_x$	$t_y$	$t_z$	$\vartheta_x$	$\vartheta_y$		$\vartheta_z$
Méthode itérative (Simple ITK)	1.89	Mean	1.618	1.8289	1.875	0.794	0.893	0.922	9.65
		SD	1.102	1.234	1.319	0.393	0.581	0.648	
Notre méthode (CNN)	1.67	Mean	1.556	1.695	1.739	0.684	0.706	0.776	<b>0.07</b>
		SD	1.099	1.202	1.106	0.423	0.414	0.416	

## Méthode de recalage élastique entre échographie 2D/coupe scanner X 2D par apprentissage profond avec apprentissage non-supervisé

Cette étude vise à réaliser le recalage non rigide entre les échographies 2D et le scanner X afin de prendre en compte, à une certaine phase du cycle cardiaque, de légères déformations du cœur qui résultent de la respiration du patient ou à l'insertion de la sonde. Dans ce cas (et dans un le cas plus général de l'imagerie médicale), l'acquisition d'une vérité terrain fiable est difficile, d'où l'exploration d'approches non supervisées pour le recalage d'images.

L'approche d'apprentissage non-supervisé que nous proposons est la suivante (Figure 7) :

- 1) Un réseau est utilisé pour estimer le champ vectoriel de déplacement entre une image échographique et une section issue du scanner X. Dans notre cas, l'architecture du réseau que nous utilisons est similaire à celle de U-Net composé d'une section de codage suivi d'une section de décodage avec des connexions entre elles à chaque niveau. Les étapes de codages capturent les caractéristiques hiérarchiques de la paire d'images d'entrée qui sont utilisées pour estimer le champ vectoriel de déplacement dans l'étape de décodage.
- 2) Le champ vectoriel de déplacement estimé est utilisé pour déformer l'image scanner X donnée en entrée.
- 3) L'image CT déformée est alors comparée à l'image échographique à l'aide d'une mesure de similarité (Information Mutuelle pour les raisons mentionnées dans les études précédentes).
- 4) Le réseau est entraîné en optimisant la métrique de similarité d'image (c'est-à-dire par rétropropagation de la dissimilarité) en utilisant l'optimiseur Adam. Après l'entraînement, le réseau peut être appliqué pour le recalage de paires d'images non vues.

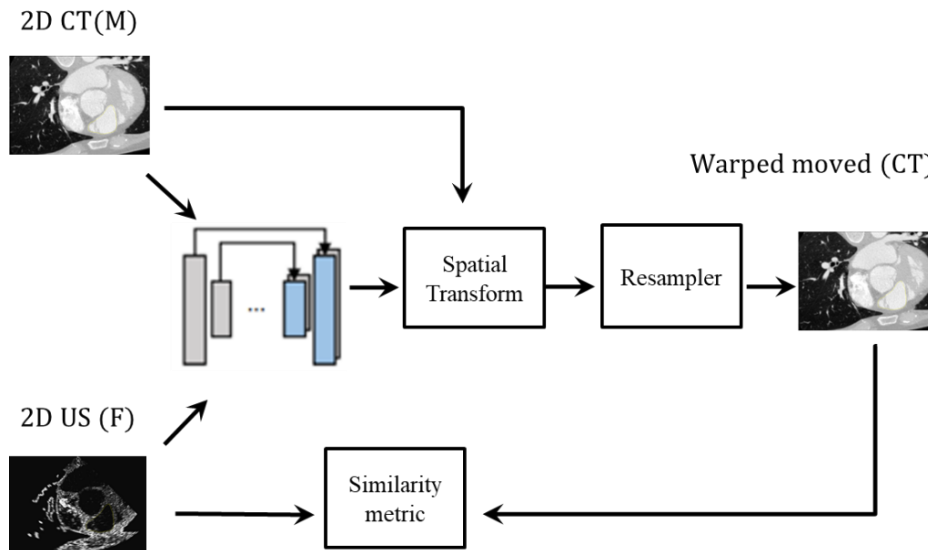


Figure 7 – Schéma de l'approche pour l'apprentissage de notre modèle.

L'apprentissage et l'évaluation ont été menés sur des paires d'images Scanner X et images échographiques simulées. Dans un premier temps, nous avons extrait arbitrairement 250 coupes du

volume CT dans une gamme de  $\pm 5$  mm en translation et  $\pm 5^\circ$  en rotation autour d'une pose initiale située dans la ligne centrale de l'œsophage au milieu du volume. Pour chacune de ces coupes nous avons créé l'image échographiques correspondante en appliquant quelques déformations artificielles sur la coupe CT (déplacement aléatoire des points d'une grille éparse et interpolation spline) avant de simuler l'échographie. L'image CT déformée sert aussi de vérité terrain pour l'évaluation.

À partir de cet ensemble de données, le réseau a été entraîné en sélectionnant de manière aléatoire 175 paires d'images. Les 75 paires restantes ont été utilisées pour la validation.

Pour la validation, nous avons comparé les résultats de recalage obtenus par notre à ceux obtenus par un recalage itératif non-rigide par champ de déformation de forme libre B-Spline implémenté dans la bibliothèque SimpleElastix. Nous avons utilisé la même mesure de similarité (Information Mutuelle) dans les deux méthodes. Pour une paire d'images, chacune des deux méthodes donnait une estimée du champ de déformation. Nous avons appliqué les champs de déformations estimées à la coupe CT. Ceci nous permet de comparer ces images déformées à la vérité terrain. Pour cela, nous avons segmenté l'oreillette gauche dans les 3 images (vérité terrain et images obtenue par les méthodes). Ceci nous permet de mesurer le score de Dice et le distance Hausdorff entre l'image issue des méthodes et la vérité terrain. La Table 1 donne la moyenne (et l'écart type) du score de Dice, de la distance de Hausdorff et du temps de calcul pour la méthode classique (SimpleElastix) et notre méthode (CNN (U-Net)).

Table 2 : Performances de notre méthode (CNN (U-Net)) comparées à la méthode classique (SimpleElastix).

Method			
	<i>Dice sim. Coef.</i>	<i>Hausdorff distance (mm)</i>	<i>Comp. time (sec)</i>
SimpleElastix	0.7 (0.01)	1.7 (0.02)	65 (0.1)
CNN (U-Net)	0.8 (0.02)	1.2 (0.05)	0.7 (0.02)

## Conclusion

En conclusion, lors de ce travail de thèse nous avons apportés 3 contributions pour le guidage d'une thérapie trans-œsophagienne de la fibrillation cardiaque par Ultrasons Haute Intensité Focalisés. Dans les 3 contributions, l'objectif était de trouver la pose de la sonde de thérapie par dans le volume scanner préopératoire afin de faire le lien entre le point focal de la thérapie et la trajectoire de l'ablation planifiée dans le volume scanner, ceci en se servant de la seule information image (coupes échographiques) fournie par la sonde. Dans un premier temps, nous avons intégré un plan image supplémentaire fourni par la nouvelle dans la procédure itérative classique et nous avons relâché certaines contraintes anatomiques afin de réaliser un recalage échographie 2D/volume scanner X 3D. Dans un second temps et afin d'accélérer le temps de calcul de la méthode itérative classique nous avons proposé une solution de recalage rigide échographie 2D/ coupe scanner X 2D par apprentissage profond supervisé. Cette solution permettait de diminuer le temps d'un recalage

2D/2D à 3 ms (à comparer à 6 s pour la technique itérative classique) tout en préservant, voire améliorant, la précision du recalage.

Nous avons ensuite étendu cette première étude de recalage 2D/2D pour proposer un recalage 2D/3D basé sur l'apprentissage afin d'affiner l'estimation de la pose transoesophagienne de la sonde dans le volume préopératoire 3D. Le cadre proposé consistait en deux réseaux pour extraire les cartes de caractéristiques de chaque paire d'image US fixe et de sous-volume mobile CT, suivi d'une couche de concaténation, et enfin le réseau de recalage ResNet a été utilisé pour estimer les six paramètres de transformation rigide.

Comme nous l'avons montré, par rapport à une méthode itérative classique, la qualité des résultats était préservée (et améliorée dans certains cas) tandis que le temps de calcul était fortement réduit. Chaque cas d'enregistrement a pris environ 0,07 seconde (presque 140 fois moins que la méthode itérative classique).

Finalement, afin de compenser certaines déformations locales dues au mouvement respiratoire ou à la variabilité de forme entre deux phases du cycle cardiaque, nous avons proposé une solution de recalage élastique échographie 2D/coupe scanner X 2D par apprentissage profond non-supervisé. Là encore, notre solution proposait un recalage en 0,7 s (à comparer à 67 s pour la procédure de recalage élastique itérative proposée dans la librairie SimpleElastix).

Les résultats expérimentaux démontrent que les performances en termes de précision de recalage sont tout aussi bonnes, voire un peu meilleures, avec nos méthodes basées sur l'apprentissage profond qu'avec la méthode itérative classique ceci avec une vitesse de calcul beaucoup plus élevée ce qui nous permet une intégration future en routine clinique.



# Table of Contents

<b>Résumé étendu de la Thèse en français .....</b>	<b>1</b>
<b>Table of Contents .....</b>	<b>xvii</b>
<b>List of Figures .....</b>	<b>xxi</b>
<b>Introduction .....</b>	<b>1</b>
<b>Clinical context .....</b>	<b>6</b>
1.1. The heart .....	6
1.1.1. Anatomy of the heart: .....	7
1.1.2. The electrical system of the heart.....	9
1.1.3. Mechanical operation of the heart .....	12
1.2. Cardiac pathologies: cardiac arrhythmias.....	13
1.2.1. Ventricular fibrillation: Definition .....	14
1.2.2. Epidemiologic .....	15
1.3. Treatments for ventricle fibrillation .....	15
1.3.1. Defibrillation .....	16
1.3.2. Medicines .....	16
1.3.3. Implantable cardioverter defibrillator (ICD).....	17
1.3.4. Catheter ablation .....	18
1.4. Heart ablation using a transesophageal HIFU probe .....	19
1.4.1. The transesophageal approach .....	20
1.5. imaging modalities for VF therapy.....	21
1.5.1. Cardiac computed tomography (CT).....	21
1.5.2. Magnetic resonance imaging (MRI).....	22
1.5.3. Echocardiography (US) .....	23
1.6. CHORUS project.....	24
1.6.1. Objectives:.....	24
1.6.2. Prototype of the transesophageal HIFU probe .....	24
1.6.3. therapy guidance through US/CT registration .....	26
1.7. Conclusion .....	26
<b>Iterative-based Image registration: classical approach .....</b>	<b>29</b>

1.8. Introduction .....	29
1.9. Background.....	31
1.10. Slice to volume image registration .....	31
1.10.1. intensity-based image registration methods .....	33
1.10.2. Related work on slice-to-volume registration .....	38
1.11. Two 2D US-3D CT image-based Registration: our proposal framework .....	44
1.11.1. Slice extraction .....	45
1.11.2. Similarity metric.....	45
1.11.3. Optimization.....	46
1.11.4. Datasets .....	46
1.11.5. Evaluation: Experiments and results .....	48
1.12. Discussion.....	52
1.13. Conclusion .....	52
<b>Learning-Based Registration: supervised Transformation Estimation .....</b>	<b>54</b>
1.14. Introduction .....	54
1.15. Background.....	55
1.15.1. Convolutional neural network.....	56
1.15.2. Autoencoder .....	56
1.15.3. Recurrent neural network .....	57
1.15.4. Reinforcement learning.....	57
1.15.5. Generative adversarial network GAN.....	57
1.16. Related work in learning based rigid medical image registration .....	58
1.16.1. Deep Similarity based Registration.....	58
1.16.2. Supervised transformation estimation .....	59
1.16.3. 2D/3D image registration using CNN.....	59
1.17. Ultrasound to CT 2D Rigid Image Registration using CNN.....	61
1.17.1. Materials and method .....	61
1.17.2. Datasets .....	63
1.17.3. US datasets .....	65
1.17.4. Evaluation.....	67
1.17.5. Discussion .....	70
1.18. Deep learning-based for slice-to-volume image registration .....	72

1.18.1. Feature extraction .....	73
1.18.2. Concatenation and registration network.....	73
1.18.3. Datasets and implementation details .....	73
1.18.4. Network training .....	74
1.18.5. Experiments and results .....	75
1.18.6. Discussion .....	78
1.19. Conclusion .....	79
<b>Learning-Based Registration: unsupervised Transformation Estimation .....</b>	<b>81</b>
1.1. Introduction .....	81
1.2. Background.....	81
1.3. CNN in deformable medical image registration: Related work .....	82
1.3.1. Deep Similarity based Registration.....	83
1.3.2. Unsupervised Transformation Estimation.....	83
1.4. Deformable US/CT registration with a convolutional neural network.....	86
1.4.1. CNN model .....	86
1.4.2. Spatial transform .....	87
1.4.3. Loss function .....	87
1.5. Experiments and results.....	87
1.5.1. Dataset.....	88
1.5.2. Experimental protocol.....	88
1.5.3. Qualitative visual evaluation.....	88
1.5.4. Quantitative evaluation .....	89
1.6. Conclusion .....	90
<b>General conclusion and perspective .....</b>	<b>93</b>
<b>References .....</b>	<b>97</b>





# List of Figures

Figure 1 – Traitement par HIFU de la fibrillation ventriculaire par voie trans-oesophagienne.....	iv
Figure 2 – Représentation schématique de la sonde avec 32 anneaux de thérapie deux barrettes d'éléments dual mode permettant d'acquérir deux plans de vue perpendiculaires en échographie. ....	v
Figure 3 – Principe du recalage échographie 2D/scanner X 3D développé dans le cadre du projet CardioUSgHIFU [2]. ....	vi
Figure 4 – Schéma de principe du recalage de 2 plans échographiques perpendiculaires et un volume scanner X 3D.....	viii
Figure 5 – Procédure de recalage rigide échographie 2D (IF) / coupes scanner X (IM). Un réseau siamois extrait des caractéristiques des images et un réseau de recalage estime directement les paramètres géométriques de la transformation rigide (T). ...	ix
Figure 6 – Schéma de l'approche pour l'apprentissage de notre modèle. ....	xiii
Figure 1.1 – Position of the Heart in the thorax . ....	7
Figure 1.2 – Presentation of the heart anatomy. ....	8
Figure 1.3 – Presentation of the veins (blue) and coronary (red) arteries. ....	9
Figure 1.4 – Schematic of the electrocardiogram and cardiac conduction system. The pink color represents the electrical system, orange the activation of the atria, and green the activation of the ventricles. ....	11
Figure 1.5 – Comparison of healthy and pathological ECGs. (a) Normal rhythm. (b) Ventricular tachycardia (VT). (c) ventricular extrasystole (VES). (d) Ventricular Fibrillations (VF). ....	13
Figure 1.6 – Position of the paddle electrodes during defibrillation/cardioversion, position of the heart, and flow of intrathoracic energy during delivery of the electric shock. ..	16
Figure 1.7 – Implantable cardioverter-defibrillator.....	17
Figure 1.8 – Trans esophageal image-guided HIFU for minimally invasive thermal ablation in the heart.....	21
Figure 1.9 – Computed tomography. (a) Spiral computed tomography system: the table moves during acquisition while the pair of emitter-detectors rotate around it. (b) X-ray projections used in computed reconstruction.....	22
Figure 1.10 – The prototype: (a) schematic view of the prototype; (b) photography of the probe head; (c) geometrical characteristics of the HIFU transducer. All the rings are the equal area. ( $\emptyset_{HIFU}$ : diameter of the HIFU transducer; THIFU: truncation; $\emptyset_{hole}$ : diameter of the hole for the imaging probe) [4]. ....	25
Figure 1.11 – Schematic representation of the probe configuration (here with 32 rings) with the designation of the reference axes. The cross-shaped elements in the middle of the probe are the dual-mode imaging/therapeutic elements. ....	26
Figure 1.12 – 3D visualization of the position of the US slices estimated by our method inside the preoperative CT acquisition. ....	26

Figure 2.1 – The basic components of a typical registration framework are two input images, a transform, a metric, an interpolator, and an optimizer. .... 31

Figure 2.2 – Example of one of the main applications requiring slice-to-volume registration. (a) Pre-operative 3D CT and intra-operative US image, (b) After slice-to-volume registration, the 2D US and the corresponding slice from 3D CT [10]. .... 40

Figure 2.3 – The general framework of their approach. During the preoperative stage, CT/MRI datasets are reformatted following the esophagus topology. During the intraoperative stage, an intensity-based registration centered in the esophagus center is performed between the US image and reformatted CT images obtained in the previous stage. .... 43

Figure 2.4 – General framework of the registration process. .... 45

Figure 2.5 – Thorax from the superior vena cava to the stomach. Axial (a), Sagittal (b) and Coronal (c) views of the CT dataset. .... 46

Figure 2.6 – An example of the extracted CT slice and the corresponding simulated US images. (a) axial CT slice, (b) sagittal CT slice, (c)  $(x_i, y_i)$  simulated US image, (d)  $y_i, z_i$  simulated US image. .... 47

Figure 2.7 – Boxplots of the translation error between the estimated parameters and GT along each axis. In blue the errors using 2 planes and in green the errors using only one plan. .... 49

Figure 2.8 – Boxplots of the angular errors between the estimated rotations and GT. In blue the errors using 2 planes and in green the errors using only one. .... 50

Figure 2.9 – angular distance between estimated rotations and GT. .... 51

Figure 2.10 – Box plots of the mean Target Registration Error (mTRE). .... 51

Figure 2.11 – Visualization result. On (a) and (c), the 2 perpendicular simulated US images to be registered. On (b) and (d) the corresponding CT planes estimated by the registration. The US images are superimposed on these CT slices. .... 52

Figure 3.1 – Example of Convolutional Neural Network (CNN) architecture ..... 56

Figure 3.2 – GAN network example in image registration, (a) Generator Network; (b) Discriminator network [102]. .... 58

Figure 3.3 – The overall of the proposed framework. .... 62

Figure 3.4 – Architecture of the regression network. .... 63

Figure 3.5 – Volumes from MMWHS2017 datasets. .... 64

Figure 3.6 – Thorax from the superior vena cava to the stomach. Axial (a), Sagittal (b) and Coronal (c) views extracted from two CT volumes. .... 64

Figure 3.7 – Dataset’s creation workflow. .... 65

Figure 3.8 – Simulation of a CT/US image pair: a) 2D CT slice extracted from a volume with the pose and field of view (yellow) of the US probe; b) the simulated US image. .... 66

Figure 3.9 – An example of the extracted CT slices and the corresponding simulated US images in the 3D CT volume. .... 66

Figure 3.10 – Three examples of the extracted CT slices and the corresponding simulated US images slices. .... 67

Figure 3.11 – Box plots of a) the Translation Estimation Errors, b) the Rotation Estimation Errors..... 68

Figure 3.12 – the position of the fiducial points marked in pink. .... 68

Figure 3.13 – Box plots of the Target Registration Errors. .... 69

Figure 3.14 – An example of the registration of an image pair a) CT image, b) US image before registration. The overlap between the moving CT image and the fixed US image (yellow image) c) before registration, and d) after registration with the proposed method. .... 70

Figure 3.15 – Our proposal framework. .... 72

Figure 3.16 – Comparison of the error estimation for estimated rigid parameters ( $R_x$ ,  $R_y$ ,  $R_z$ ), and ( $T_x$ ,  $T_y$ ,  $T_z$ ) for our proposed method (Figures (a) and (c)) and the classical iterative approach presented by [77]. (Figures (b) and (d)). .... 76

Figure 3.17 – examples of the registration results. .... 78

Figure 4.1 – An example of the U-Net architecture. .... 82

Figure 4.2 – The general framework of the proposal approach. .... 86

Figure 4.3 – (a) 2D CT slices, (b) simulated 2D US slices with superimposed boundaries of the left atrium (yellow). The deformed moving images obtained by (c) the free-form deformation field method of SimpleElastix and (d) the proposed approach. .... 89



# List of Tables

Table 1: Performances de notre méthode (CNN) comparées à la méthode classique (SimpleITK).....	xii
Table 2 : Performances de notre méthode (CNN (U-Net)) comparées à la méthode classique (SimpleElastix).....	<b>Error! Bookmark not defined.</b>
Table 2-1 – Values of parameters used in the simulation of US images. $\lambda$ is the acoustic wavelength. ....	48
Table 2-2 – Values of some input probe parameters.....	48
Table 3-1 – Performance comparison of our proposed framework and the classical iterative method.....	77
Table 4-1– Average Dice similarity coefficient, Hausdorff distance and computation time results for SimpleElastix and CNN (U-Net) across the segmented left atrium in the US fixed image, and the segmentation mapped from the warped registered CT (see Figure 4.3). ....	90



# Introduction

Cardiovascular diseases are the first cause of death in Europe, representing nearly 35% of deaths [8]. Although the morbidity related to these diseases has been decreasing since 2013 in developed countries, thanks to the development of new therapies, thus the number of hospitalizations has been increased [8]. In the European Union, the cost of these pathologies is estimated at nearly 210 billion euros in 2015, they are representing almost 20% of total health expenditure. In general, the organization World Health Organization (WHO) estimates that cardiovascular diseases are the first leading cause of mortality in the world, with nearly 17.5 million victims in 2012, and a forecast of 23.6 million in 2030. The main identified causes of these diseases are hypertension, obesity and alcoholism [8], [9].

Heart failure (HF) is a condition in which the heart is no longer able to provide blood flow to meet the needs of the different body systems. Approximately 6.2 million people in the United States suffered from HF in 2016, and they are increasing [9]. HF can be caused by many cardiac pathologies, mainly arrhythmias, contraction asynchrony, and post-infarction scarring. The most common type of arrhythmia is atrial fibrillation, which affects around 5.3 million people in United States, and was responsible for 450,000 hospitalizations in 2014 [9].

Cardiac electrophysiology is the branch of cardiology that deals with disorders of the heart's electricity. The two main disorders are arrhythmias, when the heart no longer has a regular beat or normal rhythm, and contraction asynchrony, which is a delay in contraction between an atrium and a ventricle, between the two ventricles, or between several segments of a ventricle. All of these disorders ultimately lead to HF, or to the death of the patient in the case of ventricular fibrillation VF.

For diagnosed patients, the first steps are to improve their lifestyle, by reducing alcohol consumption, tobacco, and changing their diet. If the problems persist, drug treatment is recommended. However, this treatment may not be sufficient and other therapies involving surgical procedures may be required. The gold standard surgical technique for treating fibrillation is currently catheterized ablation and more precisely radiofrequency ablation [10].

However, the efficiency of catheter ablation is limited, estimated at just over 60% for patients with paroxysmal atrial fibrillation AF. It decreases to less than 30% for those with persistent AF. This technique is invasive. In addition, some complications related to the radiofrequency energy delivery have been reported, such as phrenic nerve injury, left atrial/ventricle-esophageal fistula and blood clots formation [10].

An alternative procedure has been proposed to avoid the problems associated to catheter ablation approaches: the myocardial ablation using HIFU delivered from the esophagus [11], [12]. This is a promising procedure for the minimally-invasive treatment of atria/ventricle fibrillation. This transesophageal technique allows ablation to be performed using an epicardial approach, without the need for surgery. HIFU technology



can be used to create thermal lesions in deep tissues, without damaging intervening tissues. It is a mini-invasive treatment that places the HIFU transducer close to the ablation zone, navigating via the esophagus. The ultrasound probe is placed close enough to the ablation area to obtain a satisfactory acoustic window. HIFU waves are delivered from the esophagus to the atrial/ventricle wall minimizing the risk of damage to the esophagus and neighboring organs.

This work of this thesis was part of a project funded by the ANR: the CHORUS project (ANR 17-CE190017). This project is a collaboration between academic laboratories with expertise in ultrasound technologies for imaging and therapy (LabTAU, Lyon), cardiology/real-time MRI technology (Liryc, Bordeaux) and image processing for treatment planning (LTSI, Rennes), while the company Vermon is a world leader in the design of ultrasound probes. Our objectives inside the CHORUS project were: 1) to process the anatomical information provided by preoperative imaging (MRI or CT) to undertake the planning of the intervention and 2) to process the intraoperative ultrasound (US) images to guide the therapy according to the planned strategy.

During therapy, the practitioner will use the 2D US images, acquired by the imaging transducer inserted in the center of the therapy probe for optimal positioning of the HIFU transducer along the esophagus with respect to area of the ventricle/atrial wall area to be treated and to verify the absence of obstacles in the firing line. Intraoperative ultrasound imaging has many advantages, it is non-ionizing, portable, low cost and fast enough to capture tissue deformation. However, the information contained in the image is relatively low and its field of view is relatively limited (to a 2D sector) and is user dependent. The registration of the preoperative high-resolution information with the intraoperative US images is a key factor in image-guided interventions. Moreover, it allows the transfer of planning information provided by the clinician to the US-guided intervention and thus reduces the user's dependence on the interpretation of intraoperative US images.

This Thesis will provide various solutions concerning the planning and guidance of transesophageal HIFU ventricle fibrillation therapy. In order to describe our contributions, we have divided this manuscript into four chapters, as follows:

**Chapter 1** introduces the clinical context and the objectives of this work. This chapter begins with the description of the functioning of the cardiovascular system, and the different cardiac disorders. We focused on the VF monitoring of the different existing therapies and their limitations, leading to the promising procedure of minimally invasive transesophageal HIFU ablation. Then the different imaging modality used for this technology are described. We emphasize the properties of the images used for treatment planning and t guidance. We also present a more detailed summary of the ANR CHORUS project and its objectives, and we focused into the new transesophageal (TEE) probe design with two perpendicular image planes instead of one as previously. Finally, we describe the scientific objectives of this thesis, namely, to propose different solutions for the registration between the 2D ultrasound images (US), and the 3D computed tomography volume (CT) used to establish the intervention planning. More specifically, we try to estimate the pose (position and orientation) of the ultrasound

images in the 3D CT volume. These solutions will be evaluated in terms of both accuracy and computational efficiency.

**Chapter 2** first presents the methodological background of the classical iterative intensity-based image registration methods. It also reviews the previous work carried out in the laboratory on this subject [7], and shows their limitations. Then the overall workflow of the registration framework is presented.

This chapter then describes our **first scientific contribution**: the two planes 2D US/3D CT rigid image registration, as well as the clinical data collected for its validation. The results of this study have been presented in the following international and national conferences:

Dahman B., Dillenseger, J.-L., "High-Intensity Focused Ultrasound (HIFU) Therapy Guidance System by Image-Based Registration for Patients With Cardiac Fibrillation", *Computing in Cardiology*, 46, Singapore, 2019, doi:10.22489/CinC.2019.315.

Dahman B., Dillenseger, J.-L., "Transesophageal HIFU cardiac fibrillation therapy guidance by 2 two perpendicular US images", *Surgetica 2019*, Rennes, 2019

Dahman B., Dillenseger, J.-L., "Ultrasound guidance of a transesophageal HIFU therapy", *RITS 2019*, Tours.

**Chapter 3** presents the exploitation of supervised deep learning-based registration methods in the medical image domain. Indeed, it has been proved that these deep learning-based approaches have outperformed the classical image and iterative optimization-based registration approaches in terms of both accuracy and computation time efficiency.

This chapter is divided into three parts, first we summarize the latest development in deep learning based medical image registration, and we mainly focus on rigid registration which is the most suitable approach for slice-to-volume image registration.

Second, we present our **second contribution**: an Ultrasound to CT 2D Rigid Image Registration framework using CNN. The preliminary results of this study were presented in an international conference:

Batoul Dahman, F. Bessier, Jean-Louis Dillenseger, "Ultrasound to CT rigid image registration using CNN for the HIFU treatment of heart arrhythmias," *Proc. SPIE 12034, Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling*, 1203414 (4 April 2022).

Finally, we present an extension of this work to a 2D US/3D CT image registration framework. We believe that this study is able to fill the research gap in real-time 2D-US and 3D-CT fusion for Ventricular Fibrillation ablation therapy guidance.

**Chapter 4** focuses on unsupervised learning-based approaches for multimodal image registration. The motivation for this chapter stems from the desire to address the challenging nature of achieving reliable ground truth acquisition in real data, also to take into account all the real conditions in the real time operation, like the heart movement and the patient's respiration and movement.

We decided to propose a framework for an unsupervised deformable image registration approach. We start this by a literature review of using unsupervised deep learning approach in medical image registration domain and their applications.

We present then, our **third contribution**: a deformable US/CT unsupervised deep learning-based registration approach. The results of this study were presented in an international conference:

Dahman B., Dillenseger, J.-L., “Deformable US/CT Image Registration with a Convolutional Neural Network for Cardiac Arrhythmia Therapy”, IEEE-EMBS, Montreal, 2020.



# Chapitre 1

## Clinical context

The heart is a complex system. An electrical network brings the heart chambers to contract in synchrony, allowing proper circulation of blood throughout the body. When this system is broken, the heartbeat becomes irregular, and the contractions out of sync, putting the rest of the functions at risk. Cardiac electrophysiology is the field of cardiology studying these disorders and their remedies. In this chapter, we first deal with the functioning of the cardiovascular system, and more particularly the anatomy and the functions of the heart (section 1.1). Different heart disorders are then mentioned in section 1.2 followed by a description of existing therapies to treat them and of their limitations (section 1.3), and the context of the work: High Intensity Focus Ultrasound (HIFU) for cardiac surgery in section 1.4. Then, the imaging modalities and descriptors considered in this work are presented in section 1.5. The work of this thesis is part of ANR project and follows some previous work that will be presented in section 1.6, and we also present in this section the objectives set out in these works. Finally, the conclusion of this chapter (section 1.7).

### 1.1. The heart

The function of the cardiovascular system is to deliver oxygen ( $O_2$ ) and nutrients to all systems of the human body using blood as a transport vector. It is made up of two main elements, the circulatory system, the network used to connect the systems, and the heart, the pump allowing to circulate blood in this network.

The heart is an important component of the cardiovascular system that helps circulate blood to the organs, tissues, and cells of the body. Blood travels through blood vessels and is circulated along pulmonary and systemic circuits [13]. The human heart is located in the thoracic cage, in a space called the mediastinum, which is located between the two lungs [14], between the upper thoracic opening and the diaphragm, and between the sternum and vertebral bodies as shown in Figure 1.1. It is a hollow muscular organ that is somewhat pyramid shaped made up of two independent parts, working in synchrony to set the blood in motion in the circulatory system.

The circulatory system is made up of two parallel circuits: the general circulation, connecting the heart and the various  $O_2$  consuming systems (muscles, brain, etc.), and the pulmonary circulation, connecting the lungs and the heart. For each of these circuits, the blood is put into movement by a specific half of the heart, the right and left chambers. These circuits are made up of blood vessels called an artery if they leave the heart towards a system, and vein if they leave a system towards of the heart.

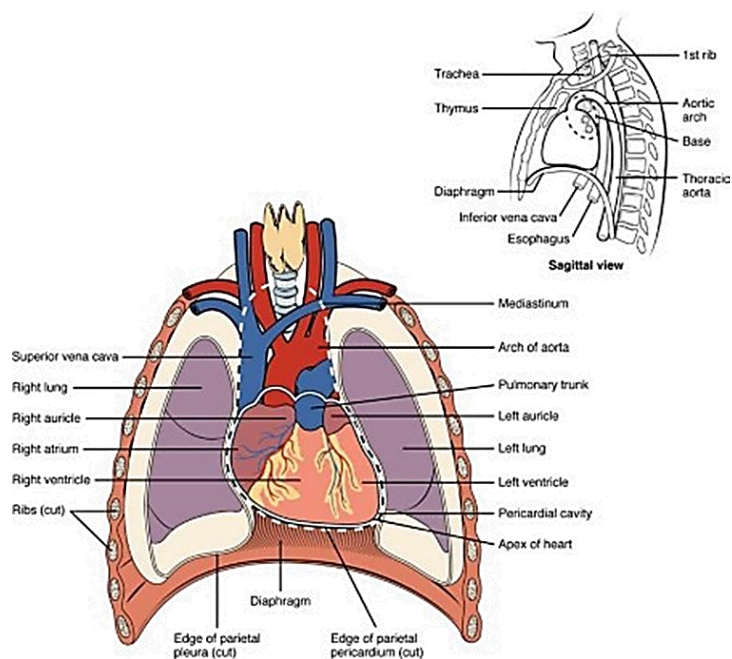


Figure 1.1– Position of the Heart in the thorax.

### 1.1.1. Anatomy of the heart:

The human heart is divided into two halves. Each half consists of a larger pumping chamber (ventricle) and a smaller filling chamber (atrium). The atrium and the ventricle are separated by cardiac tissue, the atrioventricular septum, each of these halves is supplied with blood through a vein (vena and pulmonary for right and left heart respectively) and expels blood through an artery (pulmonary and aorta for right and left heart respectively). The blood is admitted through the atrium and expelled through the ventricle. These cavities are surrounded by contractile tissue, the myocardium, which expels the blood they contain. In order to avoid any backflow and maximize the efficiency of the pump function performed by the heart, non-return valves are placed at the entrance and exit of the ventricle: the heart valves. All the elements presented in this section can be seen in Figure 1.2.

**The atriums:** The atriums are the cavities that receive blood from the veins, and whose contraction allows the transfer of blood to the ventricles. The right atrium (RA) and the left atrium (LA) have similar pseudo-spherical shapes. They are separated by the interatrial septum [15].

**The ventricles:** The ventricles are the chambers that receive blood from the atria, allowing by their contraction the transfer of blood to the pulmonary and aortic arteries for the ventricle right (RV) and left ventricle (LV) respectively. The RV and the LV have conical shapes. They are separated by the interventricular septum [16].

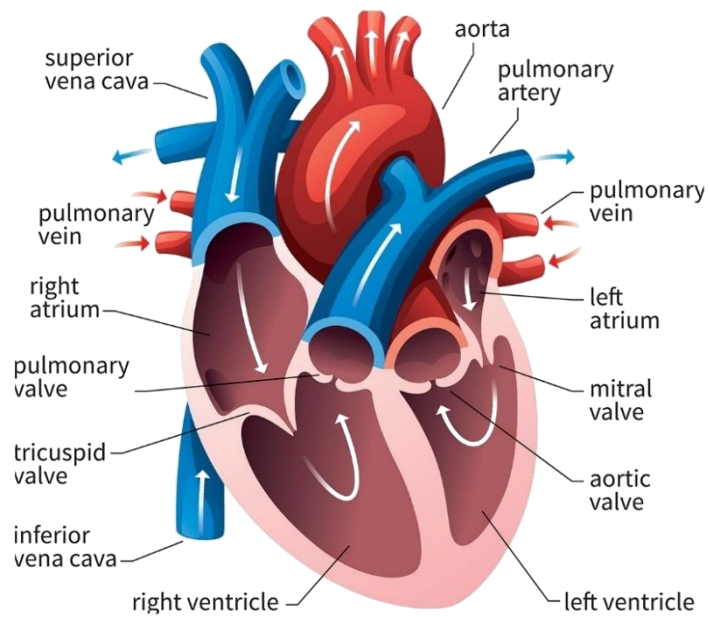


Figure 1.2 – Presentation of the heart anatomy.

**The myocardium:** The myocardium is the muscle covering each of the cavities, and whose contraction allows blood to be expelled from them. The myocardium is very thin around the atria, thicker around the ventricles. An important difference exists between the thickness of the myocardium of the RV and of the LV, for the benefit of the LV. This is explained by the fact that the route of general circulation is much longer than that of the pulmonary circulation. The myocardium is made up of muscle cells called cardiomyocytes. These are cells specific muscles of the heart that are distinguished by different electrical and mechanical properties. They are:

- **Intetanisables** After contraction, cells cannot be re-excited immediately. They are therefore incapable of prolonged contractions.
- **Conductive** The cells transmit to the neighboring cells the excitement that made them contract.

**Heart valves** at the entrance and exit of each ventricle, there is a valve allowing the direction of blood flow from the atria to the ventricles (valves atrioventricular), and from the ventricles to the arteries (sigmoid valves). The four valves are:

1. The **tricuspid** valve, separating the RA from the RV.
2. The **pulmonary** valve, separating the RV from the pulmonary artery.
3. The **mitral** valve, separating the LA from the LV.
4. The **aortic** valve, separating the LV and the aorta.

The mitral and tricuspid valves (atrioventricular valves) are supported by the attachment of fibrous cords (chordae tendineae) to the free edges of the valve cusps. The chordae tendineae are, in turn, attached to papillary muscles, located on the interior surface of the ventricles these muscles contract during ventricular systole to prevent prolapse of the valve leaflets into the atria. There are five papillary muscles in total.

Three are located in the right ventricle and support the tricuspid valve. The remaining two are located within the left ventricle, and act on the mitral valve.

**Blood irrigation** Being a muscle, the myocardium needs to be supplied with blood in order to be supplied, among other things, with  $O_2$ . Arteries from the aorta circulate on the surface of the myocardium and supply it with oxygenated blood. Symmetrically, veins emerge from the myocardium, and meet in a vein called the coronary sinus, which directly joins the RA. We name these vessels are called the coronary vessels, as they form a crown around the heart [17]. A diagram of these vessels is visible in Figure 1.3.

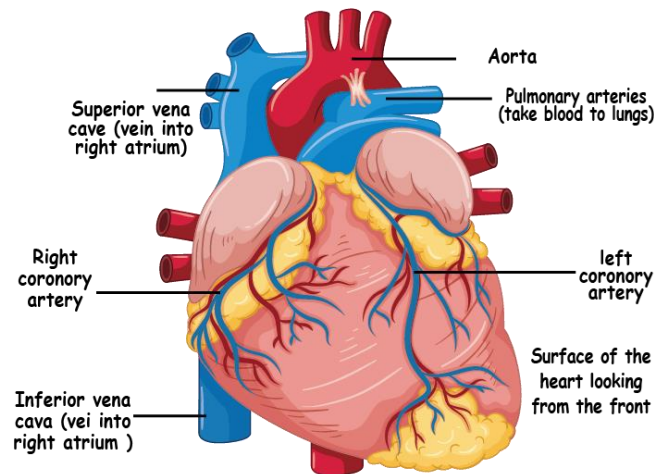


Figure 1.3– Presentation of the veins (blue) and coronary (red) arteries.

If the part of the myocardium is no longer supplied by a coronary artery, the muscle becomes necrotic. It is a myocardial infarction. Tissues are then replaced by tissue scar made up of collagen fibers, called fibrosis.

### 1.1.2. The electrical system of the heart

The contraction of the myocardium is controlled by an electrical impulse generated by specific tissues of the myocardium. This impulse is then transmitted by electrical conduction in the rest of the heart. The frequency of these impulses is regulated by the nervous system.

**Nodal tissue:** In the previous paragraph, we talked about muscle cells that make up the myocardium: cardiomyocytes. There are several categories, depending on their function. In particular, some are self-excitable, allowing them to periodically generate and autonomously an electrical impulse of a few millivolts. These cardiomyocytes make up the nodal tissue, consisting of two nodes and a branched filament [18].

1. **The sinus node:** located at the junction between the right atrium and the superior vena cava.
2. **The atrioventricular node:** located between the ostium of the coronary sinus and the tricuspid valve.
3. **The bundle of His:** located in the prolongation of the atrioventricular node, descends along the interatrial septum, then into the interventricular septum.

The two nodes are connected by three thin bundles, called internodal bundles. All these elements are visible in Figure 1.4.



### **1.1.2.1. Electrical stimulation of the whole heart**

In normal cases, electrical impulses stimulating the myocardium are generated by cardiomyocytes of the sinus node. We then speak of normal sinus rhythm. The natural frequency of sinus node depolarization is 60 to 80 beats per minute. It should be noted that the heart has safety devices, allowing at the atrioventricular node to take over the function of the sinus node in the event of dysfunction, by generating the impulses itself. However, it is not as effective, and does produce only 40 to 50 beats per minute. It is the same for the bundle of His, with an independent activity of about 30 beats per minute.

The impulse propagates through the two atria via the three internodal bundles, and through the cardiomyocytes step by step. Around the atrioventricular valves (mitral and tricuspid valves) are two fibrous rings (mitral and tricuspid) that block the flow of electricity. The activation front therefore has only one exit point: the atrioventricular node. Connected to the atrioventricular node, the bundle of His transmits the impulse to the two ventricles. It is composed of two branches, left and right, corresponding to the two ventricles. Each branch leads the pulse through the interventricular septum to the apex ventricle associated with it, before moving up along its side wall. At the end of branches, the Purkinje fibers finish propagating the stimulation. The conduction speeds are variable along the nodal tissue. In the atrioventricular node, conduction is slow to give the atria sufficient time to contract completely, in order to fill the ventricles as much as possible. Conversely, conduction is extremely fast in the His bundle and the Purkinje lattice in order to synchronize the contraction of both ventricles and maximize their pumping function.

### **1.1.2.2. Heart rate regulation**

The heart rate is regulated by two nervous systems. The parasympathetic system helps to reduce the natural frequency of the sinus node through the Vagus nerve, secreting acetylcholine. Conversely, the Sympathetic system helps increase this frequency by releasing adrenaline and norepinephrine. These two nerve impulses make it possible to modulate the autonomic activity of the sinus node, and to adapt the heart rate to the physiological need of the rest of the body [19].

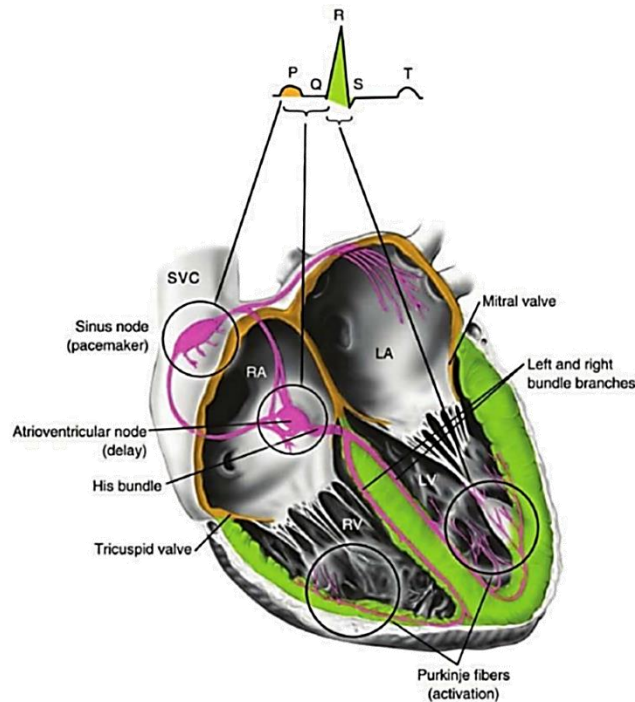


Figure 1.4 – Schematic of the electrocardiogram and cardiac conduction system. The pink color represents the electrical system, orange the activation of the atria, and green the activation of the ventricles.

### 1.1.2.3. Surface electrocardiogram

Cardiomyocytes produce an electric field as they contract. The surface electrocardiogram (ECG) surface is a clinical routine examination, consisting in measuring these fields using electrodes placed on the surface of the patient's skin, at a sampling frequency of approximately 15 kHz [20]. The electrical signal detected is in the order of a millivolt. It is a quick, painless and non-invasive examination, which highlights various abnormalities cardiac conditions (Cardiac pathologies). The ECG is acquired by 10 electrodes, from which 12 curves, called leads are calculated. Four electrodes are placed at the ends of the four limbs (ankles and wrists), and six on the patient's chest. They are separated into two groups of 6 derivations: the peripheral (or frontal) leads, and precordial leads.

**Correspondence between ECG and electrical activity:** A normal ECG is shown schematically in Figure 1.4. Depending on the derivations, this profile varies. We distinguish here different morphologies that correspond to particular electrical activities of the heart during the cardiac cycle [21]. These morphologies are:

1. **P wave:** Corresponds to the depolarization of the atria.
2. **QRS wave:** Also called QRS complex, corresponds to the depolarization of the ventricles.
3. **T wave:** Corresponds to the relaxation of the ventricles. The “atrial” T wave is masked by the QRS complex.

The PR interval (onset of the P wave at the onset of the QRS complex) corresponds to the time of transmission of the impulse from the sinus node to the bundle of His via the atrioventricular node. The PR segment (end of the P wave at the start of the QRS

complex) is included in the PR interval and corresponds to the transmission time of the electrical impulse in the atrioventricular node. Likewise, the QT interval (start of Q wave to end of T wave) corresponds to the set polarization / repolarization of the ventricles, *i.e.*, the full contraction time of the ventricles. In the remainder of this manuscript, a cardiac cycle is considered the RR cycle, *i.e.*, the interval between two peaks R.

### 1.1.3. Mechanical operation of the heart

The two heart pumps operate in synchronicity in a cyclical fashion. Fibers cardiac organs form oriented bundles, determining the movements of the myocardium during its contraction [22].

**The cardiac cycle:** In the previous paragraphs, we have seen the different elements allowing blood to circulate (cavities and myocardium), to direct blood flow (valves), and to trigger a heartbeat (nodal tissue). Every beat breaks down into four phases, which form the cardiac cycle.

The phases of contraction of the myocardium are called systole contraction. The relaxation phases of the myocardium are called diastole expansion. The four moments of the cycle are:

1. **Ventricular filling** (diastole) As a result of the electrical impulse generated by the sinus node, the atria contract, causing the ventricles to fill. The flow of blood moving to the ventricles decreases as the difference in pressure between atrium and ventricle decreases.
2. **Isovolumetric contraction** (systole) Once the depolarization front has reached the bundle of His, the ventricles contract. The pressure there increases, until exceed atrial pressure, causing the atrioventricular valves to close, and further increases until the blood pressure exceeds, causing to open of the sigmoid valves. This isovolumetric contraction is noted because the volume of blood contained in the ventricles remains constant.
3. **Isotonic contraction** (systole) Muscle contraction continues, driving out the volume of blood contained in the ventricles. As with ventricular filling, the flow of expelled blood decreases as the pressure difference between the ventricles and the arteries shrink. The pressure in the ventricles decreases until it becomes less than blood pressure, causing the sigmoid valves to close. This contraction is known as isotonic because the myocardial tension is constant throughout this phase.
4. **Isovolumetric relaxation** (diastole) During this very short phase, the myocardium becomes release causing the pressure in the ventricles to decrease, until the pressure becomes lower than the pressure in the atria. Therefore, the atrioventricular valves open, and a new ventricular filling occurs.

We define different volumes of the heart:

- **End-diastolic volume** (EDV): the volume of blood contained in the ventricle when it is fully released (maximum volume).

- **End-systolic volume (ESV):** the volume of blood contained in the ventricle when it is fully contracted (minimum volume).
- **Stroke volume (SV):** given by the difference between the volume end-diastolic and end-systolic, which is the volume of ejected blood.

## 1.2. Cardiac pathologies: cardiac arrhythmias

Rhythm disturbances, or arrhythmias, correspond to disturbances in the frequency and heartbeat, they are related to a dysfunction of the electrical conduction pathway in the cardiac tissue [11]. Some of these arrhythmias are mild or chronic, and are caused by temporary factors (stress, fatigue...). Others, on the contrary, can be disabling in the everyday life, leading for example to a heart failure (HF) for example, or even leading to the death of the patient. They demand the diagnosis and treatment by a specialist.

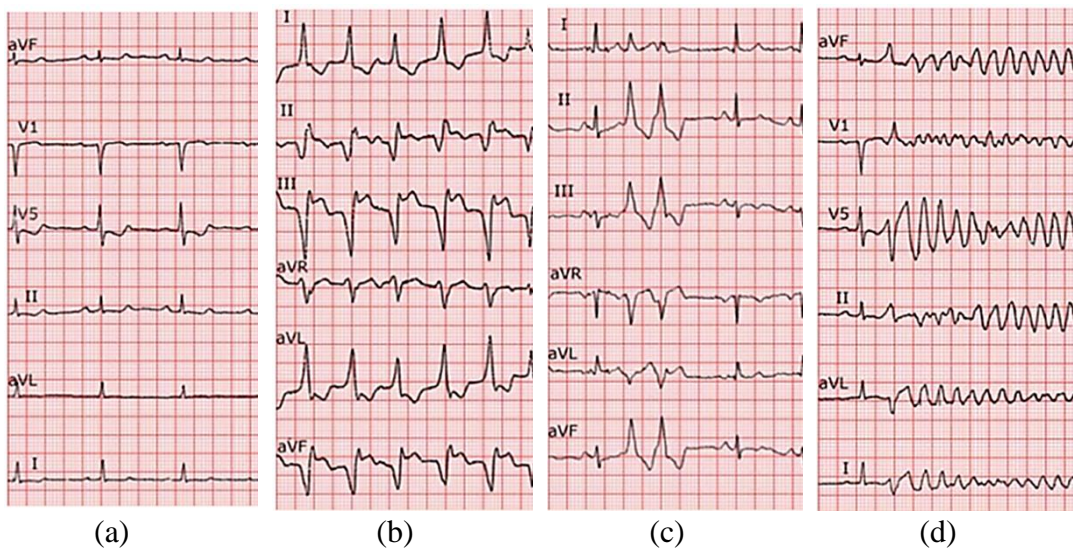


Figure 1.5 – Comparison of healthy and pathological ECGs. (a) Normal rhythm. (b) Ventricular tachycardia (VT). (c) ventricular extrasystole (VES). (d) Ventricular Fibrillations (VF).

**Ventricular tachycardia (VT):** Tachycardia is a regular very high frequency heartbeat, greater than 180 beats per minute, without the making any special efforts from the patient. This high rate leads to HF because of the blood cannot no longer being pumped out of the heart. An ECG of a patient with ventricular tachycardia is visible Figure 1.5b. We can see on this ECG that the QRS complexes succeed each other until they touch each other.

In most cases, tachycardia is due to what we call a reentry on scar. “Reentry” refers to the reentry of the depolarization front contracting the cardiomyocytes.

In section 1.1.3, it was said that cardiomyocytes are untabulatable, due to the fact that they have a period of unexcitability after contraction. The scar areas contain regions fibrosis, unexcitable, blocking electrical conduction. Around of these, areas of surviving myocardium have speeds of very weak conduction and can form channels between cicatricial areas (isthmuses). The propagation of the depolarization front can be so slow in this channel, that the cardiomyocytes located at the end of it can be excitable again when the depolarization front reaches them. So, an echo of the initial

forehead is spread, creating a spontaneous contraction of the ventricle, before re-entering the isthmus which caused it, repeating this cycle.

**Ventricular extrasystole (VES):** A premature ventricular systole (VES) is a supplemental systole which happened in the cardiac cycle. It is observed on an ECG by the appearance of an additional QRS complex additional. It can be compared to a hiccup of the heart.

An ECG from a patient with VES can be seen in Figure 1.5c, with a cycle polluted by two VESs. VSE can occur in healthy patients, and can be facilitated by factors such as stress, anxiety, or taking a stimulating substance such as alcohol or coffee. In patients who have had a myocardial infarction, repeated VSEs may be increased risk of sudden death, such as triggering ventricular fibrillation. In the case where scar tissue is present, the mechanism causing VES is identical to that of tachycardia except that the excitation loop is not maintained. This is because the propagation of the depolarization front in the isthmus in the input-output direction is blocked by the propagation of the depolarization front in the output-input direction.

**Fibrillations:** Fibrillation is caused by an “electrical storm” that brings the cavities to contract at a very high frequency and in a disorderly manner. We distinguish here the atrial fibrillation and ventricular fibrillation, pertaining to the atria and ventricles, respectively. Atrial fibrillation is the most common type of arrhythmia. It causes degradation function of the atria, leading to heart failure (HF), and stroke risk. Arrhythmia in the ventricles may lead to ventricular tachycardia and ventricular fibrillation that result in sudden cardiac death. Ventricular fibrillation on the other hand is hemodynamically inefficient and leads to death of the patient. Therefore, it is necessary to use a defibrillator as soon as possible, to shock the heart and stop the arrhythmia attack.

### 1.2.1. Ventricular fibrillation: Definition

Ventricular fibrillation (V-fib) is one of the most dangerous types of arrhythmias, or irregular heartbeat, it is the most common arrhythmia underlying sudden cardiac death that affects the heart’s ventricles.

The V-fib occurs when the electrical signals that tell the heart muscle to pump cause the ventricles to quiver (fibrillate) instead. The quivering means that the blood is not pumping blood out to the body. For some people, V-fib may happen several times a day. This is called an “electrical storm.” Because sustained V-fib can lead to cardiac arrest and death, it requires immediate medical attention.

The ECG of ventricular fibrillation caused by following an extrasystole can be seen in Figure 1.5d. It can be seen that QRS complexes have completely disappeared compared with the normal heart rhythm (Figure 1.5a) and were replaced by a completely anarchic electrical activity.

The cause of ventricular fibrillation is not always known but it can occur during certain medical conditions. V-fib most commonly occurs during an acute heart attack or shortly thereafter. When heart muscle does not get enough blood flow, it can become electrically unstable and cause dangerous heart rhythms. A heart that has been damaged

by a heart attack or other heart muscle damage is vulnerable to V-fib. Other causes include electrolyte abnormalities such as low potassium, certain medicines, and certain genetic diseases that affect the heart's ion channels or electrical conduction.

**The most common risk factors are:**

- A weakened heart muscle (cardiomyopathy).
- An acute or prior heart attack.
- Genetic disease such as long or short QT syndrome, brugada disease, or hypertrophic cardiomyopathy.
- Certain medicine that effect heart function.
- Electrolyte abnormalities.

**The symptoms of ventricular fibrillation:**

- Near fainting or transient dizziness.
- Fainting.
- Acute shortness of breath.
- Cardiac arrest.

**To diagnose ventricle fibrillation several healthcare providers should be considered:**

- Vital signs, such as the blood pressure and pulse.
- Tests of heart function, such as an electrocardiogram.
- The overall health and medical history.
- A description of symptoms that the family, or a bystander provides.
- A physical exam.

**1.2.2. Epidemiologic**

The rate of sudden cardiac deaths (SCD) per year in Europe is ranged from 200 000 to 350 000, it can be estimated that 50–70% of the deaths are due to tachy-arrhythmic mechanisms [23]. Ventricular fibrillation (VF) precipitated by ventricular tachycardia (VT) is a common mechanism of cardiac arrest leading to SCD [24]. Despite revolutionary progress in the last three decades in the treatment of ventricular tachyarrhythmia with the use of implantable cardioverter defibrillator (ICD) therapy, it remains a major public health burden [25]. Catheter ablation (CA) is an efficient option in patients with ICD and structural heart disease, for reducing VT recurrence, and ICD shocks [26]. It is also indicated in patients without apparent structural heart disease, for treating several forms of VT. Success rate of CA varies from 35% to 75%, depending on the underlying cardiomyopathy.

**1.3. Treatments for ventricle fibrillation**

Ventricular fibrillation requires emergency medical treatment to prevent sudden cardiac death. The goal of emergency treatment is to restore blood flow as quickly as possible to prevent organ and brain damage. Treatments for ventricular fibrillation includes: defibrillation, medication, implantable cardioverters and ablation.

### 1.3.1. Defibrillation

This treatment is also called cardioversion. An automated external defibrillator (AED) delivers shocks through the chest wall to the heart. It can help restore a normal heart rhythm.

Successful defibrillation largely depends on two key factors: the duration of the VF and the metabolic condition of the myocardium. The VF waveform usually begins with a relatively high amplitude and frequency; it then degenerates to a smaller and smaller amplitude until, after approximately 15 minutes, asystole is reached, possibly because of depletion of the heart's energy reserves. Unfortunately, VF waveform measures do not appear to be useful for differentiating ischemic from nonischemic cardiac arrest etiology [27].

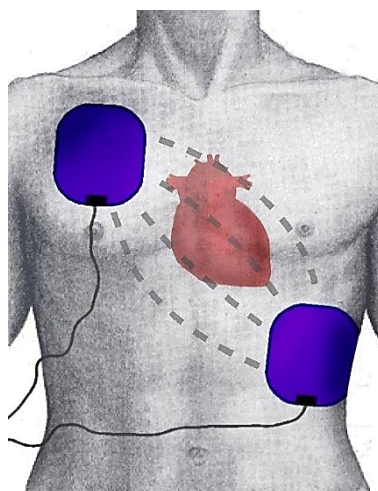


Figure 1.6 – Position of the paddle electrodes during defibrillation/cardioversion, position of the heart, and flow of intrathoracic energy during delivery of the electric shock.

Defibrillation success rates decrease about 5%-10% for each minute after the onset of VF. In strictly monitored settings where defibrillation was performed most promptly, success rates of 85% have been reported.

The goal is to use the minimum amount of energy required to overcome the threshold of defibrillation. Excessive energy can cause myocardial injury and arrhythmias.

Larger paddles result in lower impedance, which allows the use of lower-energy shocks. Approximate optimal sizes are 8-12.5 cm (3.15-4.92 inches) for an adult, 8-10 cm (3.15-3.94 inches) for a child, and 4.5-5 cm (1.77-1.97) inches for a baby. One paddle is positioned below the outer half of the right clavicle and the other one over the cardiac apex (V4 -V5) (see Figure 1.6).

### 1.3.2. Medicines

In acute ventricular fibrillation (VF), drugs (*e.g.*, vasopressin, epinephrine, amiodarone) are used to control the heart rhythm after those three defibrillation attempts are performed to restore normal rhythm. Amiodarone can also be used on a long-term basis in patients who refuse placement of an implantable cardioverter-defibrillator (ICD) or who are not candidates for an ICD. However, amiodarone has not

been shown to be of value for primary prevention of VF in patients with a depressed left ventricular (LV) ejection fraction (LVEF).

In an analysis of the association between rearrest and intraresuscitation antiarrhythmic drugs in relation to the Resuscitation Outcomes Consortium (ROC) amiodarone, lidocaine, and placebo (ALPS) trial, investigators did not find a difference in rearrest rates between those who received amiodarone or lidocaine and those who received placebo. [28] However, the electrocardiographic waveform characteristics were associated with the treatment group and rearrest and rearrest was associated with poor survival and neurologic outcomes.

### 1.3.3. Implantable cardioverter defibrillator (ICD)

An implantable cardioverter-defibrillator (ICD) is a small battery-powered device implanted near the left collarbone during a minor surgery. One or more flexible, insulated wires (leads) from the ICD run through veins to your heart to monitor the heart rhythm and detect irregular heartbeats. as shown in Figure 1.7. An ICD can deliver electric shocks via one or more wires connected to the heart to fix an abnormal heart rhythm.

ICD is a specialized device designed to directly treat many dysrhythmias, and it is specifically designed to address ventricular arrhythmias. ICDs have revolutionized the treatment of patients at risk for sudden cardiac death due to ventricular arrhythmias. A permanent pacemaker is an implanted device that provides electrical stimuli, thereby causing cardiac contraction when intrinsic myocardial electrical activity is inappropriately slow or absent.

Acute surgical complications include pain, bleeding, pneumothorax, hemothorax cardiac perforation with or without pericardial effusion and tamponade, Pulseless electrical activity following intraoperative defibrillation threshold testing. Also, Subacute ICD complications include Infection pocket hematoma, wound dehiscence, lead dislodgment, deep venous thrombosis, upper extremity edema, degradation of lead function.

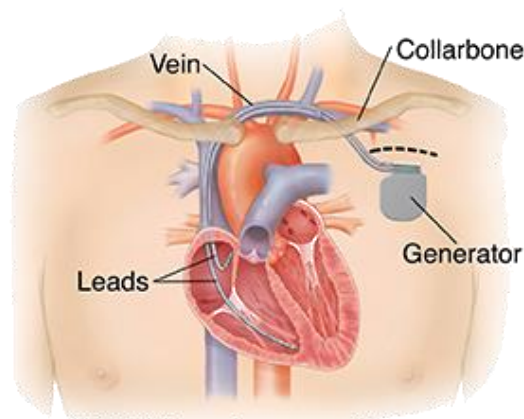


Figure 1.7– Implantable cardioverter-defibrillator.



### 1.3.4. Catheter ablation

Radiofrequency ablation (RFA) is a therapy based on the removal of tissue responsible for arrhythmias recommended for the treatment of drug-resistant arrhythmias [29]. In addition, because drug solutions (amiodarone, propafenone) have serious side effects for the patient, studies question the benefit of increasing the dosage of these treatments before resorting to ablation [30].

This therapy has recognized positive effects but is a procedure particularly long and therefore expensive.

The RF energy is a form of alternating electrical current that generates a lesion in the heart by electrical heating of the myocardium. A common form of RF ablation found in the medical environment is the electrocautery, which is used for tissue cutting and coagulation during surgical procedures. The goal of catheter ablation with RF energy is to transform electromagnetic energy into thermal energy in the tissue effectively and to destroy the arrhythmogenic tissues by heating them to a lethal temperature.

The mode of tissue heating by RF energy is resistive (electrical) heating. As electrical current passes through a resistive medium, the voltage drops, and heat is produced (similar to the heat that is created in an incandescent light bulb). The RF electrical current is typically delivered in a unipolar fashion with completion of the circuit through an indifferent electrode placed on the skin. Typically, an oscillation frequency of 500 to 750 kHz is selected. Lower frequencies are more likely to stimulate cardiac muscle and nerves, resulting in arrhythmias and pain sensation. Higher frequencies will result in tissue heating; however, in the megahertz range, the mode of energy transfer changes from electrical (resistive) heating to dielectric heating (as observed with microwave energy). With very high frequencies, conventional electrode catheters become less effective at transferring the electromagnetic energy to the tissue, and therefore complex and expensive catheter antenna designs must be used. Resistive heat production within the tissue is proportional to the RF power density, and that, in turn, is proportional to the square of the current density. When RF energy is delivered in a unipolar fashion, the current distributes radially from the source.

The current density decreases in proportion to the square of the distance from the RF electrode source. Thus, direct resistive heating of the tissue decreases proportionally with the distance from the electrode to the fourth power.

As a result, only the narrow rim of tissue in close contact with the catheter electrode (2–3 mm) is heated directly. All heating of deeper tissue layers occurs passively through heat conduction. If higher power levels are used, both the depth of direct resistive heating and the volume and radius of the virtual heat source will increase.

Although being widely accepted as a reference clinical treatment, catheter-based radiofrequency ablation mainly relies on energy deposition from the contact point between and electrode located at the catheter tip and the myocardium. However, this contact is difficult to maintain in the presence of cardiac contraction.

As a result, effective energy deposition at the target remains insufficient to ablate the pathologic tissue. Consecutive point-by-point radiofrequency procedures are often

performed to isolate the abnormal electrical pathway, which requires contiguous individual ablations that can be challenging in presence of motion. Moreover, when tissue thickness is important (e.g., typical values in the ventricle thickness are in the range of 10-15 mm), diffusion of the thermal energy from this contact point cannot guarantee achievement of a transmural lesion, which is mandatory to achieve a complete treatment.

Current procedures are mainly performed under X-Ray monitoring that provides poor information of soft tissue and mainly serves as guidance of the device to the desired cardiac chamber before 3D contact electrical mapping is performed to precisely locate the tissue to ablate. However, there is a clear lack of real-time visualization of lesion formation during the ablation, although completion of the treatment is currently assessed through variation of electrical local potentials (drop of impedance, reduction of electrical signal amplitudes). 3D navigation systems estimate the presence of a lesion using algorithms based on dosimetric studies; the area is considered destroyed when energy is applied for a sufficient period of time, with good stability and contact of the catheter to theoretically allow a proper diffusion of energy (CARTO VisiTag™ Module) [31]. Even if catheter ablation can have “acute” successful outcome, recurrence may occur after a recovery period, when the inflammatory processes surrounding the thermal lesion disappear, allowing restoration of abnormal electrical pathways and requiring expensive redo procedures.

Ways to improve the effectiveness of the procedure and reduce the risks for the patient are now proposed (e.g., heart ablation using a transesophageal HIFU probe).

#### **1.4. Heart ablation using a transesophageal HIFU probe**

High intensity focused ultrasound (HIFU) allows the creation of a thermal lesion in a tissue by focusing ultrasonic energy. The principle is as following: ultrasound beams are focused on the target tissue, and due to the significant energy deposit at the focus, the temperature within the tissue can rise to values between 65 °C to 85 °C, destroying tissues by coagulation necrosis. This modality is perfectly suited to treat deep tissues and could be designed to reach the arrhythmogenic substrate throughout the thickness of the myocardial wall.

The use of HIFU for the treatment of cardiac arrhythmia has been studied since the mid-1990s. Different approaches have been proposed: extracorporeal, extracardiac, intracardiac and transesophageal.

The extracorporeal and extracardiac approaches of [32] and [33] respectively, proved that the therapy is made difficult by the limited acoustic window due to the presence of ribs and lungs.

In contrast, intracardiac approaches place the HIFU transducer in the cavities of the heart, close to ablation zone, thus improving the acoustic window [34], [35]. HIFU catheters have been developed for pulmonary vein isolation. Steerable HIFU balloon catheters performs circumferential lesions around pulmonary veins [36]. After the first clinical trials, the success rate in patients with paroxysmal AF was similar to those obtained using RF ablation. Nevertheless, the complication rate was greater [37].

Among these complications, we can mention esophageal lesions (one causing a mortal atrio-esophageal fistula) and persistent paralysis of the phrenic nerves. These complications caused clinical trials to be stopped.

Another approach, called the Epicor Cardiac Ablation System, is an epicardial treatment using a probe designed to make ablation lines without needing to be repositioned [38]. Nevertheless, it is a complex and invasive technique, requiring epicardial surgery.

In contrast, a transesophageal technique as proposed by [12]. allows the ablation to be carried out using an epicardial approach without the needing for surgery [12]. It is a mini-invasive treatment that places the HIFU transducer close to the ablation zone by navigating inside the esophagus. The energy is thus controlled from a location close enough to the ablation zone to obtain a satisfactory acoustic window.

#### **1.4.1. The transesophageal approach**

Because HIFU enables the generation of precise thermal ablations in deep-seated tissues while preserving adjacent and intervening tissues, it has the potential to be used as an ablation technique for the heart. HIFU energy has already been used to create thermal lesions in cardiac tissues.

The feasibility of a transesophageal method using HIFU energy has been considered more recently [39], [12]. In humans, the esophagus is located just behind the heart and offers an excellent acoustic window for transesophageal echocardiography as shown in Figure 1.8. Some targeted areas are located just in contact with the anterior face of the esophagus (left atrium) or are easily accessible with a left rotation or a trans-gastric positioning (left ventricle). During the same endoscopic procedure, dynamic focusing of the HIFU beam from the esophagus could allow for targeting both the heart regions that are usually accessible with epicardial approaches and those that are usually treated with endocardial approaches. Moreover, the mid-myocardium of the ventricles could also be targeted.

An endoesophageal HIFU device allows for focusing the ultrasound beam from the esophagus in such a way that the potential energy remaining after the focal point would be dissipated by the blood flow inside the heart cavities. Just as during transrectal HIFU procedures for treating prostate cancers and during the propagation of ultrasound through the rectal wall [11], cooling the endoesophageal probe protects the esophagus from thermal damage. The risk of deleterious effects is reduced compared to the HIFU balloon catheter that delivers energy from the heart outward.

The transesophageal technique consists of generating thermal ablation using a HIFU transducer placed in the esophagus. The transducer is set on a transesophageal probe that navigates inside the esophagus in order to reach the region that lies just below the posterior wall of the heart as shown in Figure 1.8. The therapy is then generated and controlled from this location.



Figure 1.8 –Trans esophageal image-guided HIFU for minimally invasive thermal ablation in the heart.

The probe navigation and transducer positioning are carried out using images provided by an US imaging probe embedded at the tip of the therapy probe. These images are very similar as this provided by a Transesophageal Echocardiography (TEE) probe. The transesophageal technique is advantageous because:

- The esophageal approach results in a good cardiac acoustic window, particularly for the left atrium/ventricle, which is the ablation zone in the therapy.
- It is less invasive than the intracardiac approach.
- An epicardial-like therapy is obtained.
- TEE probes are routinely used in cardiac interventions.
- The blood flow helps to cool the cardiac cavities. In contrast, intracardiac approaches cannot exploit this advantage and the adjacent organs can be damaged.
- A cooling balloon can be added at the head of the HIFU probe, which cools both transducer and the surrounding tissue, thus avoiding the risk of thermal lesions of the esophagus.

### **1.5. imaging modalities for VF therapy**

In clinical routine, various conventional imaging examinations may be performed. Each of them makes it possible to observe a very precise anatomical, mechanical or electrical characteristic of the heart tissue. In this section, we present very briefly the different examinations considered in this work, as well as the descriptors studied for each.

#### **1.5.1. Cardiac computed tomography (CT)**

Computed tomography (CT) is a modality based on the measurement of absorption of x-rays by tissues. In the rest of the manuscript. This is the clinical all body volume

imaging modality with the highest spatial resolution, less than a millimetre in all three axes, making it a reference method for defining the patient's anatomy.

Digital geometry processing is used to generate a three-dimensional image of the inside of an object from a large series of two-dimensional radiographic images captured around a single axis of rotation [40].

The first commercial CT was invented by Sir Godfrey Hounsfield, who gave his name to the unit of absorption coefficients or a Hounsfield (H). For instance, absorption coefficients for air, water (similar to blood) and bones are  $-1000H$ ,  $0H$  and  $1000H$ , respectively. Thus, the differences in X-ray absorption between tissues allow for different organs in the body to be distinguished. Spiral CT is the dominant type of CT scanner technology. In this type of machines, the patient is placed on a motorized table and moved during CT acquisition while the pair of X-ray emitter-detectors rotates around it, as shown in Figure 1.9. The number of X-ray projections (angular resolution) and the size of the field of view (spatial extent of the object to be imaged), are important parameters defining the CT image resolution.

The synchronization with an ECG allows the acquisition of the heart volume at a specific phase. To reconstruct a volume at a specific phase, we can either acquire only images corresponding to this phase by synchronizing the acquisition ECG (prospective reconstruction), or acquire images continuously, and sort the phases a posteriori for reconstruction (retrospective reconstruction).

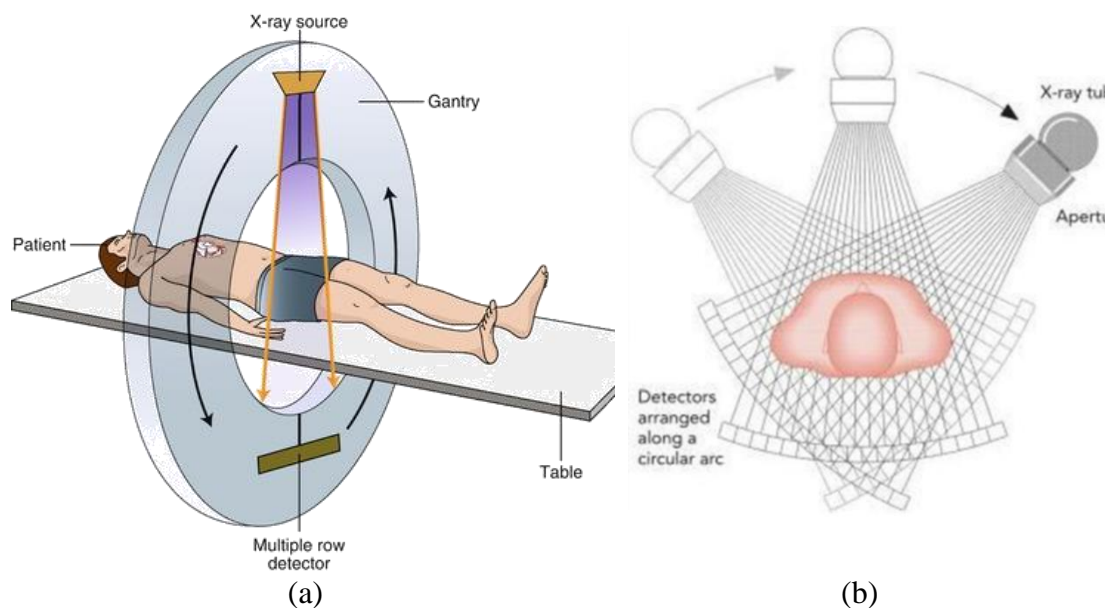


Figure 1.9 – Computed tomography. (a) Spiral computed tomography system: the table moves during acquisition while the pair of emitter-detectors rotate around it. (b) X-ray projections used in computer reconstruction.

### 1.5.2. Magnetic resonance imaging (MRI)

Magnetic resonance imaging (MRI) is based on the principle of nuclear magnetic resonance (NMR): Hydrogen protons pointing in the same direction produce a signal. The principle is to immerse the patient in a powerful magnetic field, which is disturbed by weaker fields. These disturbances change the orientation of hydrogen atoms. When

the disturbances stop, the hydrogen proton returns to the orientation induced by stable magnetic field, by emitting a measurable NMR signal. The signal is measured after a given time after the disturbances have ceased, and is broken down as follows two directions, one collinear with the magnetic field in which the patient is immersed, the other orthogonal. These components are called respectively T1 relaxations (longitudinal) and T2 relaxations (transverse). The different tissues are then distinguished by the different concentrations of hydrogen protons that constitute them.

The radiation used for MRI is approximately nine orders of magnitude smaller than in X- or  $\gamma$ -rays (used for radioisotope examinations) and is considered biologically safe.

### 1.5.3. Echocardiography (US)

Ultrasound is an imaging technique using ultrasound to detect borders between environments with different acoustic impedances. Echocardiography is the ultrasound protocol for observing the heart.

Ultrasound transducers can be integrated into different types of probes and surgical instruments to produce an echocardiography. This enables to image the heart from multiple points of view. The most commonly used US cardiac systems are:

- Transthoracic Echocardiography (TTE): the US probe is placed on the thorax; thus, the US images of the heart are acquired through the chest. This is a non-invasive system.
- Intravascular Ultrasound (IVUS): The US probe is attached on the top of a thin catheter. This system is often used to image the arteries, *i.e.*, (for navigation purposes in a stent placement procedure).
- Intracardiac Echocardiography (ICE): The US probe is placed inside a catheter. The insertion of the catheter usually begins in the femoral vein, the right internal jugular vein, or the left subclavian vein. The catheter is then guided through the vena cava into the heart cavities.
- Transesophageal Echocardiography (TEE): US transducers are placed in the head of a probe designed to be inserted into the esophagus (Figure 1.8). Compared to transthoracic echo, TEE provides a unique access to some cardiac structures including the aorta, the heart valves and the atria [41]. Moreover, TEE images are of higher quality because the transducer is located close to the heart, thus avoiding interferences from fat, lungs, and ribs.

Beside this anatomical imaging capabilities, new ultrasound-based imaging techniques offer striking perspectives. Active (Project CardioUSgHIFU ANR 2011) or passive elastography [42] permits to see the development of the thermal lesion within the myocardium. More impressively, elastographic data can be used to follow the electromechanical activation of the heart and eventually provide feedback on the treatment of arrhythmia [43].

## 1.6. CHORUS project

The work presented in this document has been developed as part of the ANR CHORUS project (ANR 17-CE19-0017).

The objective of this project is to provide ventricular fibrillation therapy by ablation using a guided High Intensity Focus Ultrasound (HIFU) via the transesophageal route. the CHORUS project aims at developing a transesophageal ultrasonic probe for image-guided thermal ablation.

CHORUS gathers a very complementary consortium. Academic laboratories have expertise in ultrasound technologies for imaging and therapy (LabTAU), cardiology/real-time MRI technology (Liryc) and image processing for treatment planning (LTSI) while the company Vermon is a world leader in ultrasound probes provider.

### 1.6.1. Objectives:

Combination of a more effective and a less invasive ablation technique that would offer a real-time image-based quantitative monitoring of lesion formation could therefore improve the Effectiveness of arrhythmias ablation procedures by:

1. Developing a transesophageal ultrasound probe for image-guided thermal ablation.
2. Improving image guidance for real-time injury assessment prediction.
3. Linking intraoperative 2D US images to high resolution preoperative anatomical 3D imaging through registration.

The work of my thesis concerns this third aspect. The preoperative anatomical 3D volume is provided by a classical cardio cine CT. The US images will be provided by the prototype of a transesophageal HIFU dual-mode probe.

### 1.6.2. Prototype of the transesophageal HIFU probe

Constancier et al [31] proposed the design of a transesophageal probe that integrates both therapy and imaging function for AF treatment using the mini-maze procedure [31]. The geometrical specifications of the probe and the evaluation of its viability to perform the mini-maze HIFU procedure were made using digital simulations. The first prototype of the transducer was a spherical truncated shape cut into several isosurface concentric rings, as shown in Figure 1.10 . This probe was used to perform cardiac ablation from the esophagus in an ex-vivo experiment on pigs.

The probes use a transducer that can be moved, which enables the focus of the ultrasound beam to be changed and for different depths to be reached.

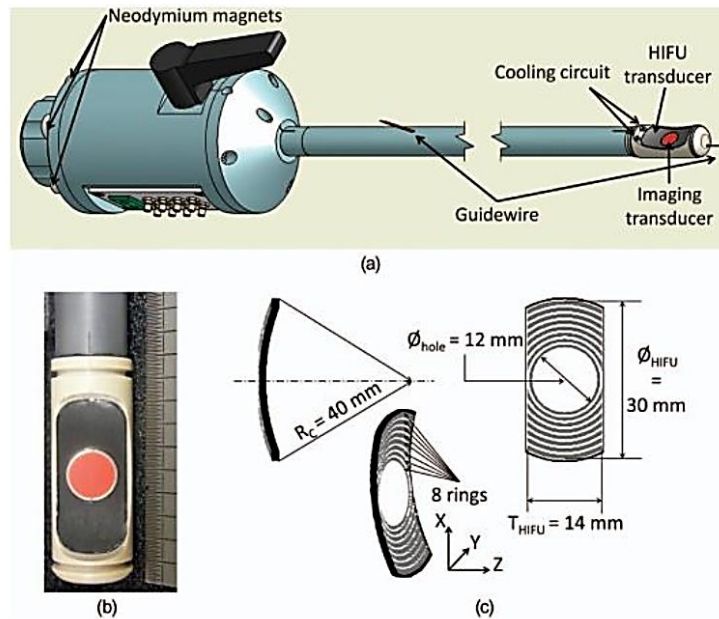


Figure 1.10 –The prototype: (a) schematic view of the prototype; (b) photography of the probe head; (c) geometrical characteristics of the HIFU transducer. All the rings are the equal area. ( $\varnothing_{HIFU}$ : diameter of the HIFU transducer;  $T_{HIFU}$ : truncation;  $\varnothing_{hole}$ : diameter of the hole for the imaging probe) [4].

As part of the CHORUS project, the partners are developing a new transesophageal probe to treat cardiac arrhythmias by HIFU. The objective of this probe is to perform HIFU lesions on pathogenic areas of the heart, primarily targeting the ventricular walls. The lesions should be capable of being made at a depth of 11 cm and should be transmural (that is, treat the entire thickness of the wall) without damaging the intervening tissues. It is also planned to include in the probe two perpendicular 2D US imaging arrays, one is perpendicular to the probe axis and the other is along the axis as shown in Figure 1.11.

This probe integrates both therapy and imaging mode. The transducer is a spherical truncated shape cut into several is surface concentric rings which used only for therapy function (the focalization of ultrasound could be realized electronically thanks to the increase of the number of rings), and two imaging arrays are placed perpendicularly to each other in the center of the global probe. These imaging arrays could be a dual-mode for the imaging and therapy functions. The implementation of a 3D imaging to visualize more accurately lesions created by the HIFU module could be a good improvement compared to the CardioUSgHIFU probe.

The active length of the probe should be 50 mm in order to increase the pressure gain, and imaging strips should be used in therapy to maximize the emitting area (and minimize secondary lesions on the skin as much as possible). The frequency of the imaging strips and therapy rings has been set at 3 MHz (Figure 1.11).



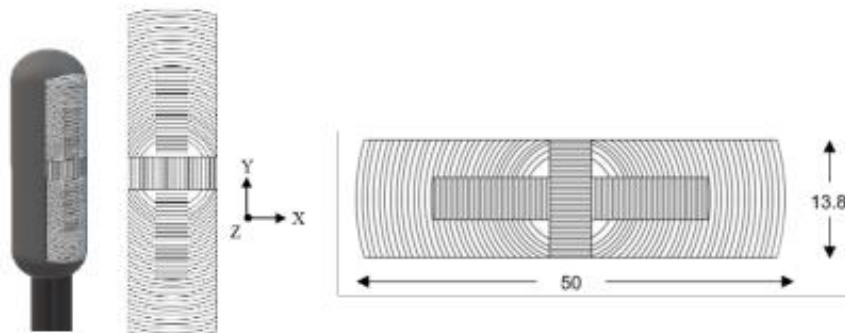


Figure 1.11 – Schematic representation of the probe configuration (here with 32 rings) with the designation of the reference axes. The cross-shaped elements in the middle of the probe are the dual-mode imaging/therapeutic elements.

### 1.6.3. therapy guidance through US/CT registration

The work of my thesis concerns the planning and guidance of therapy from ultrasound images acquired by a TEE imaging device placed in the transesophageal probe. More specifically, we are interested in registration solutions between the 2D ultrasound images (US), and the 3D computed tomography volume (CT) used to establish the intervention planning. More particularly, we try to estimate the pose (position and orientation) of the ultrasound image in the 3D CT volume as shown in Figure 1.12. And this uses only the US image information without any external localization system.

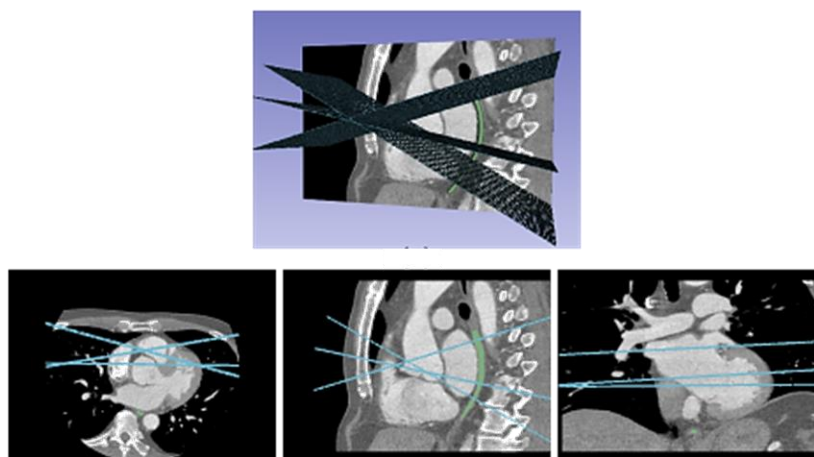


Figure 1.12 – 3D visualization of the position of the US slices estimated by our method inside the preoperative CT acquisition.

## 1.7. Conclusion

In this chapter, we have presented the clinical context of this work. First, we recalled some basics of the functioning of the cardiovascular system, and more particularly of the heart.

Various pathologies were then mentioned, in particular ventricular fibrillation FV, as well as the therapies implemented to treat them. General treatment options include

medicines, medical procedures, and lifestyle changes. If medication does not offer a solution, intracardiac catheterized Radio-frequency ablation is advised. Significant complications can be caused by this relative mini-invasive procedure: it remains still invasive, it presents some location and power control difficulties with a risk of non-transmural ablation, and also, they exist some risks of injury to external organs (*i.e.*, esophagus) [44], For these reasons, a transesophageal HIFU ablation therapy has been proposed.

Minimally invasive procedure using a US guided transesophageal HIFU probe is a promising approach to obtain transmural lesions, reducing the risk of damage of near organs. The ultrasound images acquired to guide the therapy enables the heart to be viewed in real time from the posterior zone, particularly the region of interest in VF ablation.

In order to enhance the guidance of the transesophageal HIFU cardiac ablation, the therapist needs to link the intraoperative 2D US images to the high-resolution anatomical preoperative 3D imaging (CT/MRI), in which the ablation path has been defined. It will therefore be a question of probe pose localization. This can be done by image registration that would help the therapist to adjust the HIFU focal point to the planned ablation path by providing them with the relevant geometric information directly to locate his instruments This approach is described in the next chapters.



## Chapitre 2

# Iterative-based Image registration: classical approach

### 2.1. Introduction

With the advent of different medical imaging modalities, clinicians can now perform diagnosis and therapy in a minimally-invasive manner. The fusion of the information brought by these complementary modalities is a key point in such therapies. This can be done by image registration.

The work presented in this document has been developed as part of the ANR CHORUS (ANR 17-CE19-0017) project, mentioned before. This project aimed to propose instrumentation and to carry out preliminary validation of HIFU-based ablation approaches for the ventricle fibrillation treatment. In this context, the registration of preoperative CT/MR with two perpendicular intraoperative US has been studied for image-guided interventions. This approach enables: (1) the transfer of diagnostic information and the intervention planning provided by the clinician to a US-guided intervention; (2) the reduction of user dependency on the interpretation of intraoperative US images.

This stage aims to guide HIFU VF therapy by integrating high resolution anatomical preoperative information inside 2D intraoperative US cardiac images. This requires the registration of the 2D-US images with the preoperative CT volume. This is a challenging task, because the images to be registered do not have the same spatial dimensions: the preoperative is a CT volume while the intraoperative image is a slice. We have to perform a 2D-3D registration. In our specific case, the 2D-3D registration consists of finding the pose of the TEE probe inside the 3D CT volume, in which the ablation path was defined. To do this, we have to search the best alignment between the US image slice with the corresponding sampled plane inside the 3D CT volume. Here it is necessary to define: (1) the kind of alignment (rigid or elastic transform), (2) the constraints of the search (parameters of the transformation) and (3) the measure of the alignment (metric).

Two hypotheses will be considered in the next sections:

**Hypothesis 1:** The alignment can be global (preserving the shape of the imaged objects), or elastic (modifying the shape according to local deformations). The choice of the kind of alignment is driven by the study of the movements and deformations introduced by the cardiac and respiration cycles [45]. The cardiac cycle introduces a periodic deformation of the heart. The heart can be imaged in one or several phases of the cardiac cycle using ECG-gated acquisitions. If both images to register correspond

to the same cardiac phase, the movement or deformation due to the cardiac cycle does not affect the registration, because the form of the heart is the same. In this case a rigid transform is sufficient. As the diaphragm is close to the heart, respiration introduces some deformation and movement during the respiration cycle. The effect of this respiration is very complex; indeed, it has not yet been modeled and shows high inter-subject variability [46]. A coarse approximation could be given as: (a) a displacement of the heart inside the human body due to respiration and (b) a deformation of the cardiac structures themselves due to the pressure or the interaction with other structures. The motion of the heart is predominantly in the cranial-caudal direction, with small displacements in the two other orthogonal planes and a rotation, especially at the apex of the heart [46]. In our specific imaging case, we can make the assumption that (i) the deformations caused by respiration are small in the region of interest of the therapy, (ii) either the US probe undergoes the same translational movement as the heart, or this translation can be compensated by a translation of the probe along the esophagus. Given these assumptions, we can therefore consider that a rigid transform is sufficient to describe the type of relationship between the 2D-US and 3D preoperative images. Another reason could be that we do not have enough information to obtain a confident elastic alignment because the deformation introduced by the respiration is volumetric and the US image just shows the deformation in 2D.

**Hypothesis 2:** The nature of the image information according to the modality is a special issue. For CT the value of the voxel intensity is related to the Hounsfield scale which gives relatively global and homogeneous information about each tissue. In contrast in US imaging, the information is mainly the signal reflected by the boundaries between tissues but also the speckle which is the result of the distribution of inhomogeneities within each tissue.

There are two ways to compare image information in the process of alignment: feature-based or intensity-based. Feature-based comparison needs an additional segmentation step of the objects of interest in both image modalities. This process is usually time or computational expensive and the segmentation errors are also propagated to the posterior registration and can have a direct impact on the final accuracy. For these reasons we chose rather to use an intensity-based approach which extracts the objects information directly from the gray levels without any additional process. The question is then, can the US image heterogeneous speckle information be used in an intensity-based approach? In cardiac imaging, such a US-CT intensity-based multimodal registration has been first reported by Huang [47]. The authors used Mutual Information (MI) to directly (or after a simple thresholding) compare CT and US cardiac images.

In this chapter we investigate the use of slice to volume registration in the context of transesophageal image guided intervention. We start with a comprehensive literature review about slice to volume registration of biomedical images, then we introduce the proposed framework, which aims to find the position of two perpendicular intraoperative US images into the preoperative 3D CT volume. Finally, our results were evaluated on simulated and real data in the context of two perpendicular 2D US and 3D CT registration.

## 2.2. Background

Image registration is the process of aligning and combining data coming from more than one image source into a unique coordinate system. Medical image registration seeks to find an optimal spatial transformation that best aligns the underlying anatomical structures. This problem has become one of the pillars of computer vision and medical imaging.

Let's start with a typical registration framework where its components and their interconnections are shown in Figure 2.1. The basic input data to the registration process are two images or volumes: one is defined as the *fixed* image  $I_F(x)$  and the other as the *moving* image  $I_M(x)$ , where  $x$  represents a position in N-dimensional space. The *transform* component  $\mu$  represents the spatial mapping of points from the fixed image space to points in the moving image space. The *interpolator* is used to evaluate moving image intensities at non-grid positions. The *metric* component  $\mathcal{M}(I_F, I_M \circ T)$  provides a measure of how well the fixed image is matched by the transformed moving image. This measure forms a quantitative criterion to be optimized by the *optimizer* over the search space defined by the parameters of the *transform* [48].

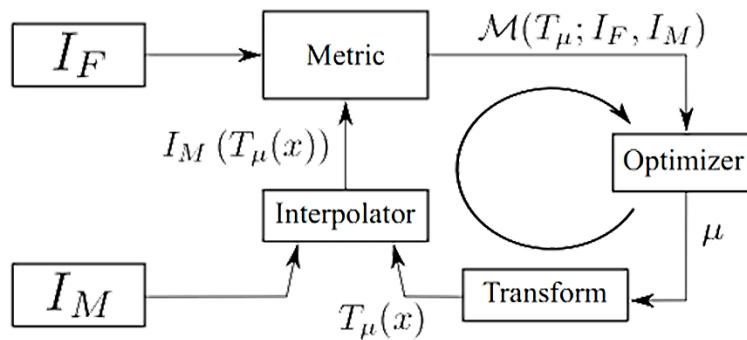


Figure 2.1 – The basic components of a typical registration framework: two input images, a transform, a metric, an interpolator, and an optimizer.

Slice-to-volume registration, a particular case of image registration problem, has received further attention in the medical imaging community during the last decade. In this case, instead of registering images with same dimension, we seek to determine the slice (corresponding to an arbitrary plane) from a given 3D volume that corresponds to an input 2D image.

## 2.3. Slice to volume image registration

Slice-to-volume registration could be considered as an extreme case of 3D-3D registration, where one of the 3D images contains only one slice, even if theoretically true 3D-3D registration methods cannot be extrapolated in a straightforward way to the slice-to-volume scenario. This holds particularly for registration methods based on image information, since the descriptors used to quantify similarities between images, normally assume that the amount of information available from both images is balanced. The fact that a single slice (or even a few sparse slices) provides less information than an entire volume, should be explicitly considered in the problem formulation. Moreover, specific geometrical constraints like planarity and in-plane

deformation restrictions, arise in the case of slice-to-volume registration, which are not applicable in the setting of dimensional correspondence.

We will start by giving a formal definition of the slice-to-volume registration. Given a 2D image  $I_F(x)$  and a 3D volume  $I_m(x)$ , we seek a mapping function  $\mu'$  that optimally aligns the tomographic slice  $I_F$  with the volumetric image  $I_m$ , through the minimization of the following objective function:

$$\mu' = \arg \max_{\mu} M(I_F, I_m; T_{\mu}) + R(\mu) \quad 2.1$$

where  $M$  represents the image similarity term (matching criterion) and  $R$  the regularization term. Note that this mapping may be rigid or non-rigid, depending on whether we allow image  $I_F$  (or its corresponding reformatted slice from  $I_m$ ) to be deformed or not. If we estimate only a rigid mapping (*i.e.*, we calculate a 6 degrees of freedom rigid transformation or even a more restrictive one), we name the problem rigid slice-to-volume registration. In some cases, we can also infer some sort of deformation model, or we consider more expressive linear transformations (such as affine transformations). We call it non-rigid registration.

The matching criterion  $M$  measures the similarity between the 2D image and its corresponding mapping (slice) to the 3D volume. Usually, it is defined using intensity information or salient structures from  $I_F$  and  $I_m$ . A complete discussion about matching criteria in the context of slice-to-volume registration will be presented in next section. The regularization term  $R$  imposes constraints on the solution that can be used to render the problem well posed. It also may encode geometric properties on the extended transformation model (plane selection or plane deformation in case of non-rigid registration). The choice of a regularizer depends on the transformation model. While models like rigid body transformations can be explicitly estimated even without regularizer, the term  $R$  becomes crucial in more complex non-rigid scenarios to ensure realistic results.

In the context of slice-to-volume registration, the regularizer can be used to impose planarity constraints to the solution (when out-of-plane deformations are not allowed) or to limit the out-of plane deformation magnitude guaranteeing realistic and plausible transformations. When available, prior knowledge about tissue elasticity can also be encoded through the regularizer. We aim at optimizing the energy defined in equation 2.1, by choosing the best  $\mu'$  that aligns the 2D and 3D images.

In our context, of the registration of preoperative CT with intraoperative US, we have considered rigid registration, as mentioned in the introduction of this chapter (hypothesis1: If both images to register correspond to the same cardiac phase, the movement deformation due to the cardiac cycle does not affect the registration, because the form of the heart is the same. In this case a rigid transform is sufficient).

Registration methods can be classified as intensity-based (the use of voxel intensities to quantify similarity), feature-based (the use of some features that could be easily detected in both images), geometric (the use of a sparse set of salient image locations

to guide the registration) or hybrid methods (which combine both intensity-based and geometric strategies). We have mentioned in our second hypotheses, that in our application the nature of the image information is a specific issue (US/CT) that has already been the subject of attempts to use intensity-based methods. We will therefore focus on this class of registration.

### 2.3.1. intensity-based image registration methods

Intensity-based registration methods are based on measurements computed directly from pixel/voxel intensities without the need of landmark identification nor segmentation.

Image registration can be monomodal when the slice and the volume are captured with the same type of image technology or multimodal when slice and volume refer to different modalities, *e.g.*, US slice and MRI or CT volume. In the first case, the task of measuring the similarity between the images is simpler, since pixel/voxel intensity values corresponding to the same anatomical structure are highly correlated, or even identical in both images.

In case of multimodality, where the relation between pixel intensities is not obvious, there are two major alternatives: to continue to use an image based matching criterion but we need then to define more complex similarity measures or to adopt a geometric or sensor-based strategy.

#### 2.3.1.1. Matching criterion

The matching criterion (also known as (dis)similarity measure, merit function or distance function) quantifies the level of alignment between the images, and it is typically used to guide the optimization process of the transformation model [49]. This criterion depends on the nature of information exploited in the matching process.

Image registration can be monomodal or multimodal.

**In the monomodal case**, metrics based on direct comparison of gray levels like Mean squares or Normalized correlation can be used:

**Mean squares** which are the mean square difference over all pixels/voxels, defined as:

$$MS(T_\mu; I_F, I_M) = \frac{1}{N} \sum_i^N (I_F(\chi_i) - \psi(I_M(T_\mu(\chi_i))))^2 \quad 2.2$$

With  $\chi_i$  a given in  $I_F$ ,  $N$  its number of voxels and  $\psi(\cdot)$  a given interpolator. The mean square metric has an ideal value of zero.

**Normalized correlation** which computes the pixel-wise cross-correlation between the intensities of the images to be registered, normalized by the square root of the autocorrelation of each image. When the two images are identical, the measure equals



one. Normalized correlation may perform better than means squares since the main tissues are clearly associated to each others.

$$NC(T_\mu; I_F, I_M) = \frac{\sum_i^N I_F(\chi_i) \cdot \psi(I_M(T_\mu(\chi_i)))}{\sqrt{\sum_i^N I_F^2(\chi_i) \cdot \sum_i^N \psi^2(I_M(T_\mu(\chi_i)))}} \quad 2.3$$

**In case of multimodality**, where the relation between pixel intensities is not obvious, some of the most challenging cases of image registration arise when images of different modalities are involved. In such cases, metrics based on the direct comparison of gray levels are no longer applicable. It has been extensively shown that metrics based on concepts derived from Information Theory, like Mutual information, Normalized Mutual information, Entropy correlation coefficient, Joint entropy, point similarity measure based on Mutual information, Energy of the histogram, Correlation ratio, Woods criterion are more efficient than others. In a previous work of our team [4], they evaluate the use of different intensity-based measures in our specific US/CT registration application in order to select the most adapted similarity measure from a set of metrics reported in the literature. They found that Mutual information and Woods criterion gave the best results.

**Mutual Information (MI)** are well suited for overcoming the difficulties of multimodality registration. The concept of Mutual information is derived from Information Theory, which measures the statistical dependence or information redundancy between the image intensities of corresponding distributions in both images, that is assumed to be maximal if the images are geometrically aligned [50]. It requires an estimation of the joint and marginal probability density functions (PDFs) of the intensities in every image. MI is defined as:

$$MI(T_\mu; I_F, I_M) = H(I_F) + H(T_\mu(I_M)) - H(I_F, T_\mu(I_M)) \quad 2.4$$

with  
the marginal  
PDF

$$H(I) = - \int pI(i) \cdot \log(pI(i)) \cdot di \quad 2.5$$

the joint PDF

$$H(I_F, I_M) = - \int p(i_F, i_M) \cdot \log(p(i_F, i_M)) \cdot di_F \cdot di_M \quad 2.6$$

Compared to the volume-to-volume scenario, the information used for the slice-to-volume registration is sparse in nature. The estimation of these marginal and joint PDFs for every slice -especially in slices of low image resolution/number of samples- is a hard task and may redound to poor MI-based registration results. One of the main drawbacks of MI, is that it varies when the overlapping area between the images changes, *i.e.* it is not invariant to changes in the overlap region throughout registration. It could happen that while estimating the transformation model, some potential

solutions lie out of the volume. In such cases, an overlap invariant function would be of choice. To this end, a modified version of MI, Normalized Mutual Information has been proposed.

**Normalized Mutual Information** (NMI), is simply the ratio of the sum of the marginal entropies and the joint entropy [51]. Another advantage of NMI with respect to MI is its range: it conveniently takes values between 0 and 1. NMI has already been used for slice-to-volume registration [52], [53].

$$NMI(T_\mu; I_F, I_M) = \frac{H(I_F) + H(T_\mu(I_M))}{H(I_F, T_\mu(I_M))} \quad 2.7$$

**Woods criterion:** it is defined as

$$WC(T_\mu; I_F, I_M) = \frac{1}{L} \sum_{i_M \in T_\mu(I_M)} L_{iF} \frac{\sigma_{iM}}{m_{iM}} \quad 2.8$$

Where:

$$\sigma_{iM} = \sqrt{\frac{1}{L_{iM}} \sum_{x \in \Omega_{iM}} I_F^2(x) - m_{iM}^2} \quad 2.9$$

where  $m_{iM}$  is the average intensity in the fixed image  $I_F$  inside the subregions  $\in \Omega_{iM} \subset T_\mu(\Omega_M)$ , *i.e.*, subregions where the image intensities in image  $I_M$  are  $i_M$ .  $L_{iM}$  is the number of voxels in subregion  $\Omega_{iM}$ . The original WC is multiplied by  $-1$  to make the optimum a maximum instead of a minimum.

### 2.3.1.2. Geometric transformation

The registration process consists in finding a spatial transformation  $T$  that maps points in the fixed image  $I_f$  to homologous points in the moving image  $I_m$ .

Geometric registration finds correspondences between meaningful anatomical locations or salient landmarks. Transformation models explain the relation between the slice and the volume being registered and are the outcome of the registration process.

One way to classify spatial transforms is to consider linear and non-linear transforms. Linear transforms are defined by a global transformation matrix, often defined by using homogeneous coordinates, and which are applied to the whole image. Non-linear transforms (also called elastic or nonrigid transformations) are defined by a set of local transforms linked together by a regularization process (*e.g.*, using splines for the free-form deformation method or smoothing for Demon's algorithm).

They are often classified according to their degrees of freedom to Rigid, Affine, Homography and deformation (B-splines or thin-plate splines) transformation.

**Rigid transformations** deal with global rotations and translations. Rigid image transformation model accounts for rotation and translation parameters. It is usually expressed as a 6 degrees of freedom (6-DOF) transformation or composed by 3 rotation and 3 translation parameters.

A rigid registration of a point  $x$  in a 3D space ( $x \in \mathbb{R}^3$ ), denoted as  $T_{\mu^R}(x) = R \cdot x + T$ , three parameters defining the translation  $T \in \mathbb{R}^3$  and three parameters for the angles of rotation ( $\theta \in \mathbb{R}^3$ ) around the axis of the coordinates system defining the rotation matrix  $R$ :  $\mu^R = (T, \theta)$ , with  $\mu^R \in \mathbb{R}^6$ . Rigid transform is also called isometric transform because it preserves distances, angles and orientation.

Such a basic model is the most common choice in literature for slice-to-volume registration. Rigid transformations are expressive enough to explain simple slice-to-volume relations. They can deal with in-plane and out-of-plane translations and rotations.

Clinical scenarios that do not inherit image distortion -like simple inter-slice motion correction or basic nature image guided surgeries [54]–[56] can be modelled with rigid transformations. When out-of-plane motion is avoided, even simpler models can be used. [57] proposed to recover in-plane slice rotations in cardiac MR series, using the stack alignment transform. In-plane translation along  $X$  and  $Y$  axes and rotation around a user-supplied center of rotation for the individual slices were parameterized independently.

Restricted rigid body transformations can be a convenient initialization component of a complete slice-to-volume registration pipeline 6-DOF rigid body transformations are part of nearly all slice-to-volume registration algorithms. Literature seeking deformable registration, often initially employs rigid alignment to account for big range displacements. The standard way to estimate 6-DOF rigid transformations, consists in minimizing an energy functional (based on an intensity based or geometric matching criterion) often with a continuous optimization algorithm (see section 2.3.1.3) where the search space is part of the Euclidean group of rigid transformations.

**Affine transformation:** An affine transformation is a linear transformation capable of modeling translation, rotation, non-isotropic scaling, and shear. 3D affine transform can be parametrized by a 12-dimensional vector:

$$\theta_{Aff} = [a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33}, t_x, t_y, t_z] \quad (2.10)$$

Such hat points  $p_B = [x_B, y_B, z_B]$  are mapped to points  $p_A = [x_A, y_A, z_A]$  according to:

$$p_A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} p_B + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (2.11)$$

**Homography transformation:** A homography transformation deforms a given quadrilateral  $Q_B = \{Q_{B1}, \dots, Q_{B4}\}$  into any other given quadrilateral  $Q_A =$

$\{Q_{A1}, \dots, Q_{A4}\}$  while keeping collinearity. It is more flexible than affine transformation, as it can handle perspective since parallel lines need not remain parallel. Homography is the model relating 2-D images (pinhole projections) of a 3-D plane.

**Deformable models** like **B-splines** can produce local in-plane and out-of-plane deformations. The richness of the deformation model is proportional to the number of parameters we need to specify. Therefore, a trade-off between the model complexity and power of the expression has to be found.

The B-spline deformable transform is designed to be used for solving deformable registration problems. This transform is equivalent to generating a deformation field where a deformation vector is assigned to every point in space. For this, the deformation vectors are estimated from the data on some control points located on a coarse grid, that is usually referred to as the B-spline grid. The deformation vectors on the other points of the field are then computed using B-spline interpolation from the deformations on the control points.

This transform does not provide functionalities for mapping vectors nor covariant vectors, only points can be mapped. The reason is that the variations of a vector under a deformable transform depend on the location of the vector in space. In other words, a vector only make sense as the relative position between two points. The B-spline deformable transform has a very large number of parameters and therefore is well suited for the numerical optimizer.

### 2.3.1.3. Optimization methods

Registration is treated as an optimization problem with the goal of finding the spatial mapping that will bring the moving image into alignment with the fixed image.

Optimization methods aim to determine the instance of the transformation model that minimizes or maximizes a function based on the matching criterion (see section 2.3.1.1). Depending on the nature of the variables being involved, those methods can be classified as continuous (deterministic) or discrete (stochastic). The continuous approaches exploit the entire space of parameters, while the discrete ones a discretized/quantized version of the admissible solutions. Both approaches can be combined.

Numerous problems in computer vision and medical imaging are inherently discrete (like semantic segmentation). However, this is not the case of slice-to-volume image registration. Most of the published methods about slice-to-volume registration adopt a continuous formulation.

Continuous optimization algorithms are generally iterative methods. They infer the best value for a set of parameters by iteratively updating them. A common gradient-based formulation for this strategy is given by:

$$\theta_{t+1} = \theta_t + \omega d_t, t = 1,2,3 \dots \quad (2.12)$$

where  $\theta$  is the vector of parameters,  $d_t$  is the search direction at iteration  $t$  and  $\omega$  is the step size or gain factor.

These methods can therefore be classified according to the derivative order they use to study the optimized function: (i) no derivative (*e.g.*, downhill-simplex method or Amoeba); (ii)

first order derivative (*e.g.*, gradient descent, regular step gradient descent, conjugate gradient; or (iii) second-order derivative (*e.g.*, Broyden-Fletcher-Goldfarb-Shanno method). The use of higher order derivatives generally improves the searching direction; however, their computational cost has to be carefully considered. The strength of gradient optimization methods is that, if the initialization is quite close to the optimum, they converge rapidly and with high precision. Their weakness is that they can converge to a local minimum if the initialization is far from the optimum. Also, their requirement of analytical derivation or numerical estimation of the energy function derivatives, reduce their applicability since it is usually complicated to estimate them.

Gradient descent is the simplest strategy in this category, where the search direction  $d_t$  is given by the negative gradient of the energy function. It refers to the standard continuous optimization method, and it has been widely applied to the problem of slice-to-volume registration [53], [58]. Conjugate gradient methods use conjugate directions instead of the local gradient to estimate  $d_t$ . Energy function with the shape of a long and narrow valley, can be optimized using fewer steps than standard gradient descent approach, resulting in faster convergence. [59], [60] have applied this strategy to estimate rigid and non-rigid slice-to-volume mapping functions, respectively.

On the other hand, stochastic optimization methods rely on randomness and re-trials to better sample the parameter space in searching for an optimum solution. The three most commonly used stochastic methods are Monte Carlo, simulated annealing, and genetic algorithm. As an example, at each iteration of the simulated annealing method, a random value is generated in order to accept or reject the new guess, even if this solution is degraded compared to the previous one. These methods may avoid being trapped in local optima, but their computational cost is generally higher than that of deterministic methods.

### 2.3.2. Related work on slice-to-volume registration

Fast and accurate 2D/3D registration plays an important role in many clinical applications. The term of 2D/3D registration is due to the dimension of the images involved in the registration process. However, this term is ambiguous since it describes two different problems depending on the technology set to capture the 2D data: it may be a projective (*e.g.*, X-ray) or a sliced/tomographic (*e.g.*, US) image. Even if both problems share similarities in terms of image dimensionality, every formulation requires a different strategy to estimate the solution.

#### *Projective 2D/3D image registration*

In this case, the 2D data is the projection of some 3D information on a projection plane. Fluoroscopic images are an example of this kind of data. There is so lack of perspective and different image geometry [61] inherent to both modalities. Moreover, a pixel in any 2D projective image does not correspond to only one voxel from the volume, but to a projection of a set of them in certain perspective. But the similarity measure needs a common geometry to perform. This dimensional correspondence can be obtained by

different methods like projection, back-projection or reconstruction [62], In other words, the information of the 3D volume is projected on a 2D image, and the registration consists of finding the projection parameters that maximize the similarity between the projected 2D image and the 2D data. The reader can refer to a comprehensive overview about projective 2D to 3D image registration in [62].

### ***Slice-to-volume registration***

In this case the dimensional support of the 2D data is a specific plane of the 3D volume. This is the case of some 2D US image recorded on the patient or a 2D MRI temperature map recorder on a patient to control the heating in a close loop manner [63]. In this case, pixels from the 2D data can be directly compared with the voxels from the 3D volume and classical similarity metrics can be directly used. The problem is now to find the right oblique plane in the 3D volume which maximizes the similarity criterion with the 2D data. So, the registration consists of finding the pose (3D location and orientation) of the 2D data that maximizes the similarity between the voxels in the oblique plane and the 2D data.

In this state of the art, we will focus on the latter case, which is how to find the pose of a 2D slice in a 3D volume.

#### **2.3.2.1. Image fusion and image guided interventions**

Several medical procedures such as image guided surgeries and therapies [64], biopsies [58], radio frequency ablation [65], tracking of particular organs [55] and minimally-invasive procedures [66], [67], [7], serve an important role in the care of patients. In this context, slice-to-volume registration brings high resolution annotated data into the operating room.

Generally, pre-operative 3D images such as computed tomography (CT) or magnetic resonance images (MRI) are acquired for diagnosis and can also be used to prepare the intervention planning. So, the 3D volumes are generally manually annotated by expert physicians (target, critical organs, margins...).

During the surgical procedure, intra-operative 2D real time images are generated using different technologies (*e.g.*, fluoroCT, US or interventional MRI slices) to help the physicist during his gesture.

The alignment of intra-operative images with pre-operative volumes augments the information that physicians have access to and allows them to navigate the volumetric annotation while performing the operation.

As an example, Figure 2.2 shows the volume and images available on our problematic. In this figure we can see: (a) slices of the preoperative 3D CT volume of the chest with heart and also the US image recorded in transesophageal manner; and in (b) the 2D US and the corresponding slice from 3D CT after slice-to-volume registration [7]. Even if intra-operative images have lower resolution and quality than the pre-operative one, the fact to see them side-by-side provides complementary information.

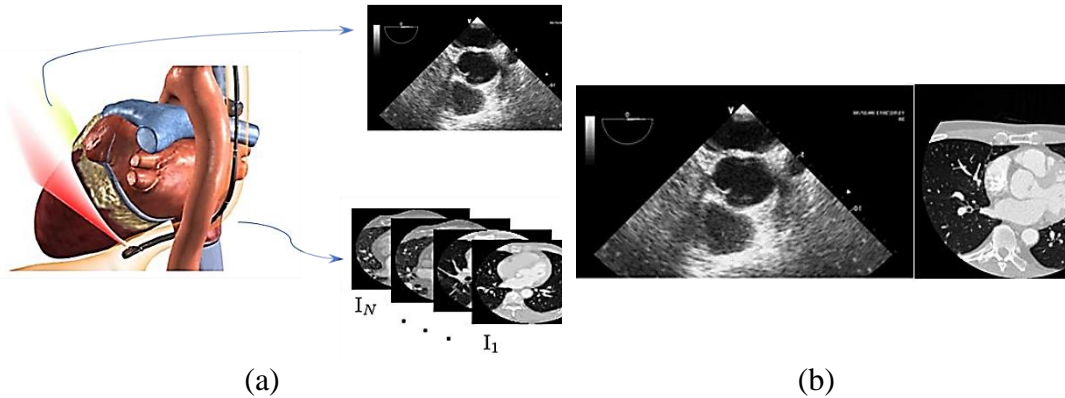


Figure 2.2 – Example of one of the main applications requiring slice-to-volume registration. (a) Pre-operative 3D CT and intra-operative US image, (b) After slice-to-volume registration, the 2D US and the corresponding slice from 3D CT [7].

A statistically significant improvement in alignment has been demonstrated when comparing automatic methods to manual (human) results, showing the importance of automatic slice-to-volume registration algorithms in the context of image fusion in image guided interventions [68]. Fei *et al.* [69] presented in a pioneer work an iconic slice-to-volume registration to the problem of image fusion in the context of image guided surgeries. The motivation was that low-resolution Single Photon Emission Computed Tomography (SPECT) can be brought to the operating room by pre-registering it with a high-resolution MRI volume, which could be subsequently fused with live time iMRI. That is how, by registering the high-resolution MR image with live-time iMRI acquisitions, Fei and coworkers could map the functional data and high-resolution anatomic information to live-time iMRI images for improved tumor targeting during thermal ablation. In [70] Birkfelner *et al* used slice-to-volume registration to fuse 2D fluoroCT with volumetric CT, which is a well-known tool for image-guided biopsies in interventional radiology. In this case, the pre-interventional diagnostic high-resolution CT with contrast agent is used to localize a lesion in the liver. However, during the intervention, the lesion is no longer visible. Thus, localizing the slice of the CT that corresponds to the intra-operative fluoroCT allows doctors to find the lesions during the biopsy. This approach only considers rigid transformations. However, interventional procedures like radio frequency ablation (RFA) or image-guided biopsies, which use fluoroCT as image guiding technology, are performed while the patient is breathing continuously. Therefore, deformations should also be taken into account when registering with the pre-operative static CT image. The influence of such deformations and the reliability of performing non-rigid registration in such scenario was discussed in [71]. It was claimed that a 2D-3D nonrigid registration solution - based on the single low quality fluoroCT- cannot be as precise as required to perform medical procedures. This is mainly due to the poor support in term of liver anatomical features provided by the fluoroCT slices. They proposed to overcome this limitation by providing an adaptive visualization [71] of the volume area surrounding the minimum estimated pose. This approach addresses the uncertainty in deformation estimation and provides more information than a single registered slice. Their method performs rigid slice-to-volume registration and includes views of the CT-Volume determined along flat directions of the out-of-plane motion parameters next to the minimum pose.

In [72] an intensity-based similarity measure is used to register interventional 2D CT-fluoroscopy to high-resolution contrast-enhanced preoperative CT image data for a radio-frequency liver ablation procedure. In [73] an intensity-based similarity metric is employed within a small region to register intra-operative 2D CT-fluoroscopy images to a preoperative CT volume to track the motion of pulmonary lesions for a robotic assisted lung biopsy. Similarly, in [74], intensity-based similarity measures are employed to register intraprocedural 2D MR images with pre-procedural 3D MR images during an MRI-guided intervention.

- **Applications using intraoperative US image registration**

Ultrasounds (US) images have been used on humans since the late 1950s or early 1960s. sound waves which is a non-ionizing radiation. Ultrasounds are particularly effective for imaging soft tissues and structures, as well as motion. In contrast, X-rays are particularly effective for imaging hard tissues or structures and air-filled parts. X-rays and ultrasounds may be used together on the same section of the body as complementary information or may be chosen one over the other depending on the gesture circumstance.

Some advantages of US against the other imaging modalities as X-ray, CT and MR are they are using non-ionizing energy, the sensor miniaturization, and the portability led them to be straightforward integrated in surgery, they are relatively low cost and provide real-time response in imaging tissue deformation. But low image resolution quality and constraints on its field of view make US to be a modality with high user dependency. Consequently, the success of interventional US imaging procedures is highly dependent on the level of experience of the practitioner.

In this context, the registration of preoperative CT/MR with intraoperative US has been studied for image-guided interventions. This approach enables: (1) the transfer of diagnostic information and the intervention planning provided by the clinician to a US-guided intervention; (2) the reduction of user dependency on the interpretation of intraoperative US images.

Laparoscopic and endoscopic interventional procedures also exploit slice-to-volume registration. The authors of [54] proposed a method to register endoscopic and laparoscopic US images with pre-operative computed tomography volumes in real time. It is based on a new phase correlation technique called LEPART accounting for rigid registration.

An intensity-based cardiac 2D US to cardiac CT image rigid registration method has been described in [75] The 2D slice is one slice imaged from a 3D object with a randomly pose and is equivalent to a slice extracted from a 3D volume of the object, their goal was to reduce the long computation time which is not suitable for real time surgeries.

The registration of preoperative CT and 2D-US was performed using an intensity-based measure in [47]. A US volume was reconstructed using acquired 2D images and electromagnetic tracking information. The preoperative CT was aligned with the US volume using tracking information. This preliminary registration is used to start an automatic intensity-based registration. MI was the metric used to obtain the parameters



of a rigid transformation. This approach was tested in an in-vivo porcine model. The registration was performed offline after all data acquisition, and the intraoperative registration time was approximately 122 s.

The approach presented [41] is a surgical navigation method using TEE. However, the registration is made with an intraoperative image. This method does not require any tracking system; thus, it can be incorporated straightforward into the operating room. The authors proposed to register and visualize 3D TEE and X-Ray fluoroscopy to guide cardiac interventions. Their method is based on the localization of the tip of the TEE probe in the fluoroscopic images. A preoperative CT image of the head of the probe with high resolution is acquired. The registration procedure iteratively repositions the CT to get different projection of the TEE probe tip, also called digitally reconstructed radiography (DRR). These images are then compared with one or multiple X-ray fluoroscopic images. The best alignment between the images results in acquiring information pertaining to the position of the TEE because the spatial transformation used to generate the projection is known. The algorithm was evaluated using a phantom and five clinical datasets. The experimental results proved that this method is viable because it is fast, reliable, and accurate.

In a previous work in our team [7], Sandoval *et al* have proposed to perform a 2D-3D (slice-volume) registration of the intraoperative 2D US and the preoperative CT/MRI without any external tracking system. More precisely the 2D-3D registration consists of finding the 3D pose (location and orientation) of the US image slice inside the preoperative 3D volume using only image-based information. They have found a way to reduce the number of degrees of freedom by using some anatomical constraints. In fact, the trajectory of the probe is constrained by the esophagus which is attached to its surrounding organs or tissues such as the vertebral column, trachea, etc. In this case, the parameters of the pose of the US probe can be simplified to i) the depth  $d$  of the US probe along the esophagus and ii) its orientation  $\theta$  around the centerline of the esophagus. The constraint which imposes the rotation around the centerline can be released by allowing a slight translation of the TEE tip position from the centerline. This choice will drive the global framework into two stages (Figure 2.3): 1) the reformatting of the preoperative CT dataset according to the esophagus topology. The main idea is to provide CT 2D slices which potentially have the same spatial location (and so the same information) as the future TEE US 2D images; 2) an intraoperative intensity-based registration between the US image and the reformatted CT 2D slices obtained in the previous stage. For this, they have proposed to find the pose of the 2D US image slice inside the preoperative 3D volume, exploiting the specific geometrical restrictions involved in this HIFU therapy with a transesophageal approach.

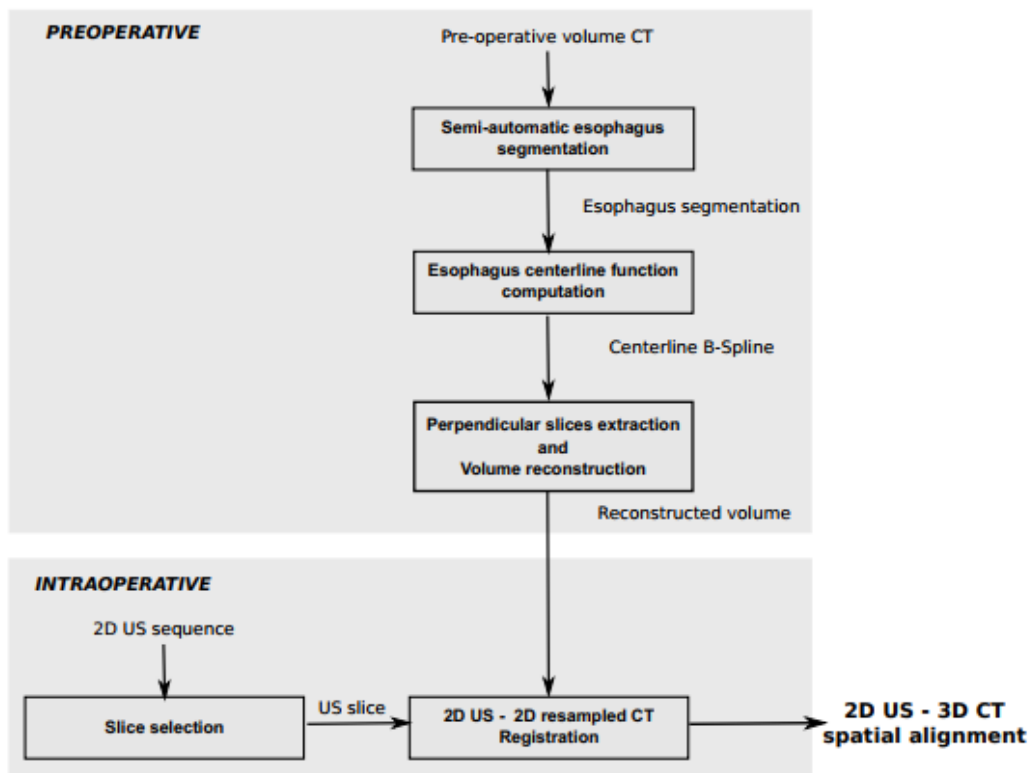


Figure 2.3 – The general framework of the 2D UD/3D CT [7]. During the preoperative stage, CT datasets are reformatted following the esophagus topology. During the intraoperative stage, an intensity-based registration centered on the esophagus axis is performed between the US image and reformatted CT images obtained in the previous stage.

The main drawback of this work is that we are not sure that the slice is really perpendicular to the esophagus axis. On other hand US image seems not to be sufficient to provide information for a precise pose estimation. For these reasons we propose to use the two perpendicular 2D US images that will be provided by the new sensor developing in the ANR CHORUS project (see section 1.4.1).

## 2.4. Two 2D US-3D CT image-based Registration: our proposal framework

In this section we present one of our contributions which consists in performing the registration of a pair of 2 perpendicular intraoperative 2D US images and a preoperative 3D CT volume using only pixel/voxel intensities information without any external tracking system (two 2D US-3D CT).

As input data, we have a pair of 2 perpendicular simulated US images similar to the real US images that will be simultaneously recorded by the dual therapy/imaging HIFU probe presented in section 1.4.3, which is still under development. As shown in Figure 1.10, in the middle of the therapy elements they are 2 perpendicular imaging strips, one perpendicular to the to the probe axis composed of 64 elements and the second along the axis composed of 2x32 elements. These elements work at an US frequency of 3 MHz.

As input, we have also a preoperative CT volume, usually a CyneCT composed of 20 volumes acquired at each 5% of the cardiac cycle.

We propose to make the registration of these 2 intraoperative spatially connected US images with the 3D preoperative CT volume. More precisely the 2D-3D registration consists of finding the 3D pose (3D location and orientation) of the geometrical linked US image slices inside the preoperative 3D volume using only image- based information.

As we mentioned before the US imaging tool is ECG-gated, we consider only the US images at the same cardiac phase as the CT and so only a 3D rigid transform with 6 DOF has to be estimated: 3 translations and 3 rotations represented by Euler angles.

Our assumption is that we have also an initial rough estimation of the pose of the probe inside the 3D CT (*e.g.*, estimated roughly by the method developed in [7] or the method described in chapter 4). From this initial pose, we will perform the proposed two 2D-3D image-based registration approaches to refine the estimation of the transesophageal probe pose.

The general framework of our registration approach is presented in the Figure 2.4. This approach is characterized by: (1) slices extraction; (2) metric; (3) optimizer.

The spatial related US images will be considered as fixed images. From a candidate 3D pose (a 3D transform) the role of the slice extraction component is to extract two-2D perpendicular CT images slices corresponding to this pose according to the geometry of the US probe. These 2D perpendicular CT images slices are considered as moving images.

The metric component provides a measure of how well the fixed images are matched by the transformed moving images. This measure forms a quantitative criterion to be optimized by the optimizer at each iteration over the search space defined by the 6 parameters of the transform  $\mu$ .

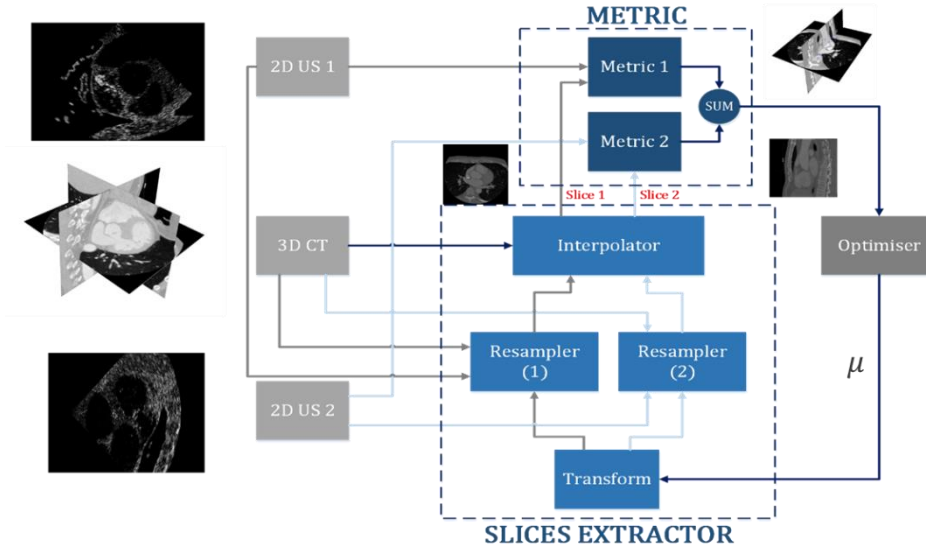


Figure 2.4 – General framework of the two 2D US/3D CT registration process.

### 2.4.1. Slice extraction

For a specific probe pose, the 3D transform allows us to define the US imaging referential system  $(\vec{O}_i, \vec{x}_i, \vec{y}_i, \vec{z}_i)$ , in which the two perpendicular planes  $(\vec{x}_i, \vec{y}_i)$  and  $(\vec{y}_i, \vec{z}_i)$  represent the spatial support of the US perpendicular slices. The CT volume is then sampled along these two planes using a resampler for each plane to provide the information in the same spatial context (same size, spatial location and orientation and sampling) as the US images.

Interpolator is required since the mapping from one space to the other will often require an evaluation of the intensity of the image at non-grid positions of the 3D volume. We have used a tri-linear interpolator, which returned value is a weighted average of the surrounding voxels, with the distance to each voxel taken as weight. Linear interpolation gives a good trade-off between reconstruction accuracy and computation complexity.

As results of the slice extraction module, we have 2 corresponding pairs of fixed/moving image, one per plane.

### 2.4.2. Similarity metric

The similarity between the corresponding pairs of fixed/moving images can now be estimated by a metric. The choice of a metric adapted to the specificity of our data is one of the most critical components of our framework. In a previous work in our team [4], the authors made an analysis the behavior of different similarity measures near of the gold standard using the framework proposed by Skerl [76]. They found that for the similarity estimation between US and CT data in the chest, the Woods Criterion and the Mutual Information have globally the best performances and should be used in the future 2D-US to 3D-CT registration. Therefore, we used Mutual Information to compare the information of the US images and the corresponding information extracted from the CT data. The global similarity will be the sum of the similarity measures obtained on the two sets of slices.

### 2.4.3. Optimization

At each iteration, the metric component  $S$  provides a measure of how well the fixed images are matched by the transformed moving images. This measure forms a quantitative criterion to be optimized over the search space defined by the parameters of the transform. We used stochastic gradient descent to estimate the pose which maximizes the global similarity.

Starting from an initial set of parameters, the optimization procedure iteratively searches for the optimal solution by evaluating the similarity at different positions inside the parameter search space [33].

The algorithm was implemented in C++ using the ITK library [77] for the similarity and optimization part, and the visualization was performed using ITKSNAP and Slicer [78].

### 2.4.4. Datasets

We evaluate our approach on preoperative 3D-CT of a patient suffering from ventricle fibrillation pathology. This CT was acquired in a clinical environment with currently available acquisition device technologies and imaging protocols. This evaluation will show the performance of our approach under real clinical conditions.

For several reasons (new probe not available, difficulties to provide an accurate ground truth transformation in real data...), we decided to work on simulated US images.

#### 2.4.4.1. Preoperative CT dataset

This study has been conducted on a CT dataset, obtained from Louis Pradel University Hospital in Lyon, France from a patient with ventricle fibrillation. An ECG-gated cardiac multislice CT image was acquired after injection of contrast agent with a Philips 64-slice scanner (Brilliance CT, Philips Healthcare) at 75% of cardiac cycle (R-R interval). The dimensions of the reconstructed volume are  $512 \times 512 \times 323$  voxels with an image spacing of  $0.546875 \times 0.546875 \times 0.55031$  mm<sup>3</sup>.

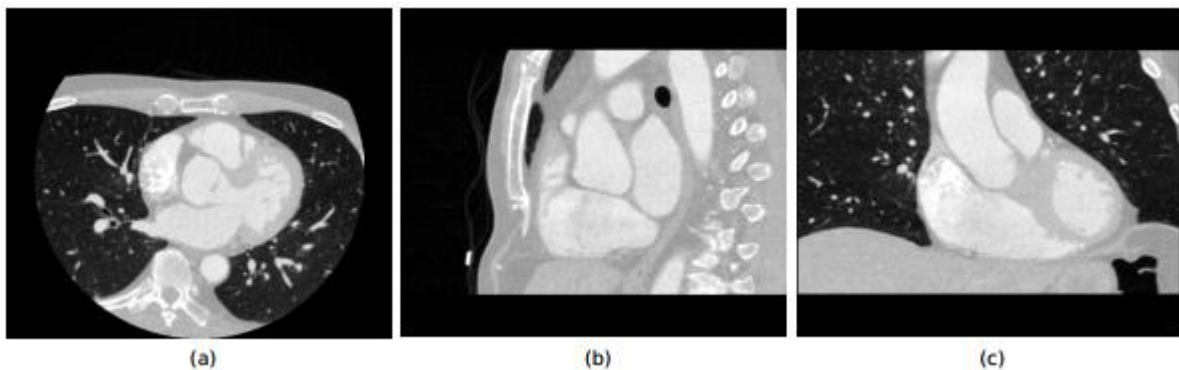


Figure 2.5 – CT of thorax from the superior vena cava to the stomach. Axial (a), Sagittal (b) and Coronal (c) views of the CT dataset

#### 2.4.4.2. Ultrasound dataset – US image simulation

Because the HIFU probe with the two perpendicular US imaging planes is still under development, we validate our method on simulated US data). First, we defined an initial pose (the ground truth-GT) inside the CT volume. From this pose we extracted two perpendicular slices from the CT (see Figure 2.6 a and b) and simulated the corresponding US slices with the method described in [3]. (See Figure 2.6 c and d).

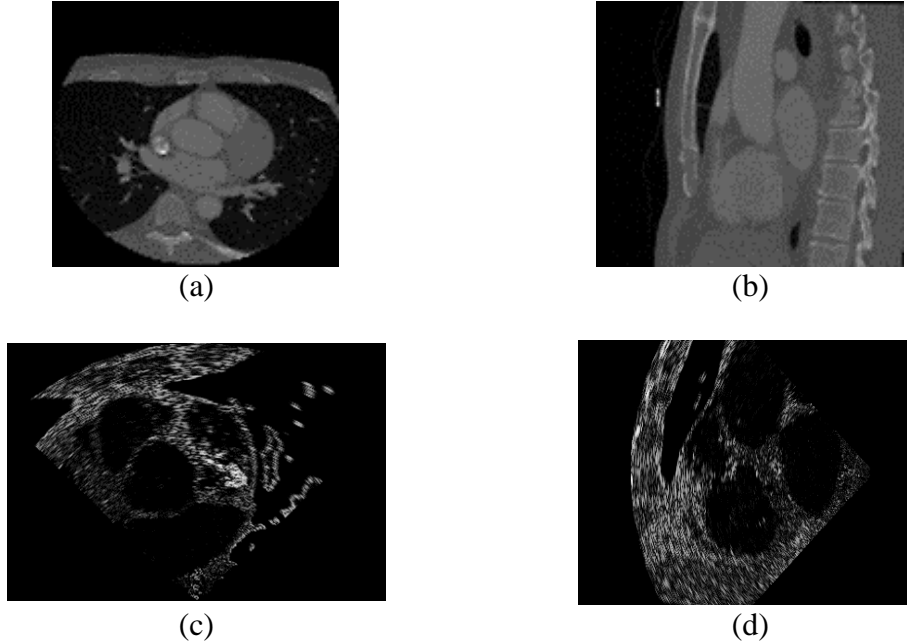


Figure 2.6 – An example of the extracted CT slice and the corresponding simulated US images. (a) axial CT slice, (b) sagittal CT slice, (c)  $(\vec{x}_l, \vec{y}_l)$  simulated US image, (d)  $(\vec{y}_l, \vec{z}_l)$  simulated US image.

In this simulator, each tissue is characterized by its acoustical impedance (see Table 2.1) and a specific spatial distribution of speckle. In this speckle model, the inter-scatterer distances are independent and randomly distributed from a gamma distribution tuned by two parameters:  $d$  which represents the mean inter-scatterer distances and so the speckle density, and  $\alpha$  a regularity parameter (see **Error! Reference source not found.**) This speckle model is able, by adjusting the speckle density and the regularity parameters, to generate the scatterers distributions analyzed in the literature like Rayleigh, Rician or  $K$  distributions.

As input of this simulator, we have also some probe parameters (see **Error! Reference source not found.**) as the US frequency, the number of elements, the curvature of the probe, the angular field of view, the depth which allows to determine the Point Spread Function (PSF) of the probe which is used when a US wave pulse interact with a scatterer.

The US radiofrequency (RF) image is obtained by the convolution of the PSF with the scatterers map. The final US image is the envelope detection of the RF image.

Using this simulator, we were able to produce 2 perpendicular US slices with a known transformation (see Figure 2.6 c and d).

Table 2.1 – Values of parameters used in the simulation of US images.  $\lambda$  is the acoustic wavelength.

	Acoustical Impedance (kg.m <sup>-2</sup> .s <sup>-1</sup> )	Speckle density (m <sup>d</sup> )	Regularity ( $\alpha$ )	Houndsfield value
Air	$4 \times 10^{-4}$	$2 \times 10^{-3} \cdot \lambda$	0.1	-1000
Water	1	$2 \times 10^{-3} \cdot \lambda$	0.1	0
Blood (contrast)	1.63	$4 \times 10^{-3} \cdot \lambda$	0.4	400
Muscle	1.65	$0.03 \cdot \lambda$	0.4	-20
Fat	1.35	$0.1 \cdot \lambda$	20	-80
Bone	$7.8 \times 10^6$	$0.02 \cdot \lambda$	20	30

Table 2.2 – Values of some input probe parameters .

US frequency	number of elements	depth	angle	curvature of the probe
7 MHz	128	150	90	3

### 2.4.5. Evaluation: Experiments and results

We arbitrarily produced 55 initial poses in a range of  $\pm 5$  mm on translation and  $\pm 5^\circ$  in rotation around the GT pose and performed the registration. The accuracy of the registration using two 2D US planes has been estimated and compared to this of the previous work with only one US plane.

For the initialization, we used some information about the endoscope (inserted length, visual analysis of the image sequence during navigation, fluoroscopy, etc.) to define a candidate zone along the esophagus centerline in which the image transducer center can be. The size of this zone corresponds broadly to 10 mm along the esophagus, which globally corresponds to the depth sampling step along the esophagus. We evaluate the influence of using two planes Vs to using a single one. The accuracy of the registration using two 2D US planes has been estimated and compared to that obtained with only one US plane using three complementary metrics.

#### 2.4.5.1. Transformation estimation error

Given a single (or multiple) slices and a volume, if the transformation  $T_{GT,i}$  that maps both images is known, we can estimate the distance between  $T_{GT,i}$  and the estimated transformation  $T_{Est,i}$ . This approach is mostly used to validate global linear transformations (see for example [63], [79] where the number of parameters to estimate is small and distances per parameter can be reported).

We can separate this transformation estimation error into 2 subsets of errors:

The translation estimation error. We chose to express it by the absolute error along each direction x, y, and z.

$$|T_{GT,i} - T_{Est,i}|_{x,y,z} \text{ (mm)} \quad (2.13)$$

and the rotation estimation error. This error is trickier to estimate since several combinations of Euler angles can describe the same 3D rotation. But any 3D rotation can also be described as a single rotation  $\theta$  around a specific rotation axes  $\mathbf{A}(A_x, A_y, A_z)$  (Rodriguez-Euler formula). Quaternion is another way to describe this rotation as a single number as:

$$q_{\bullet,i} = \cos(\theta/2) + i A_x \sin(\theta/2) + j A_y \sin(\theta/2) + k A_z \sin(\theta/2) \quad (2.14)$$

(See appendix 1). So, the rotation estimation error between the GT rotation quaternion and the estimated one can be defined as:

$$2 \cos^{-1} \left( \text{real}(q_{GT,i} * q_{Est,i}^*) \right) \text{ (degree)} \quad (2.15)$$

Figure 2.7 shows the boxplots of the translation errors along  $x, y, z$  of the 55 trials using one or 2 planes. We can see that the median translation error is reduced from 1.5 to 0.7 mm when using 2 planes. Figure 2.11 shows the rotation errors which are reduced from  $3^\circ$  to  $2.1^\circ$  when we used two perpendicular slices.

For both errors we obtained some parameters estimation accuracy improvements by adding the second plane. This improvement has been proven to be significant since we got a p-value  $< 0.032$  between all the pairs using the pairwise nonparametric Wilcoxon test [80].

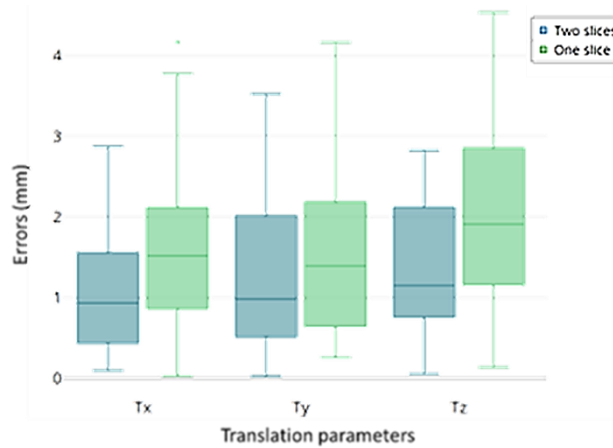


Figure 2.7 – Boxplots of the translation error between the estimated parameters and GT along each axis. In blue the errors using 2 planes and in green the errors using only one plane.



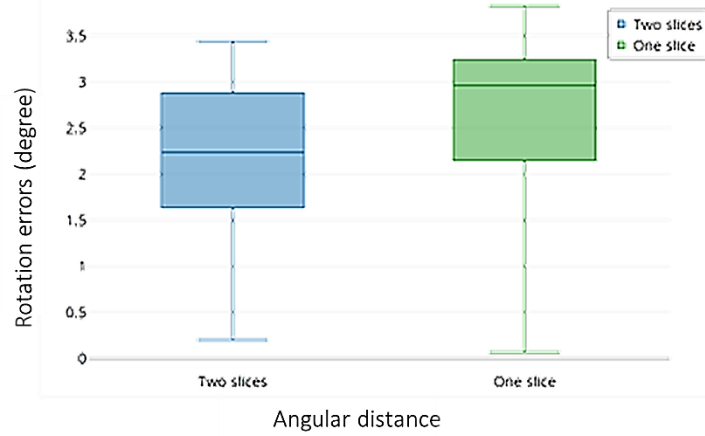


Figure 2.8 – Boxplots of the angular errors between the estimated rotations and GT. In blue the errors using 2 planes and in green the errors using only one.

### 2.4.5.2. Target Registration Error (TRE)s

Another common evaluation strategy frequently used in literature is based on the impacts of the parameter estimation error on some landmarks fitting. The idea is to annotate automatically or by an expert some points of interest which are visible in both the slices and the volume images, so that we can measure the distance between the corresponding points after registration. The distance between the ground truth and the registered anatomical landmarks is commonly referred as Target Registration Error (TRE)

In our specific application we defined eight specific feature points (or landmarks)  $P_{US,j}$  in the two 2D-US fixed images, four located in the intersection of the slices (F1, F3, F7, F8), and the others located in different places in the slices (F2, F6, F4, F5) as shown in Figure 2.9. These 8 points were then reprojected on the 3D CT volume using the ground truth transformation  $T_{GT}: P_{Est,j} = T_{GT}P_{US,j}$  and reprojected using the estimated transform  $T_{Est}: P_{Est,j} = T_{Est}P_{US,j}$ . The TRE is then defined as:

$$TER_{,j} = d(P_{GT,j}, P_{Est,j}) \quad (2.16)$$

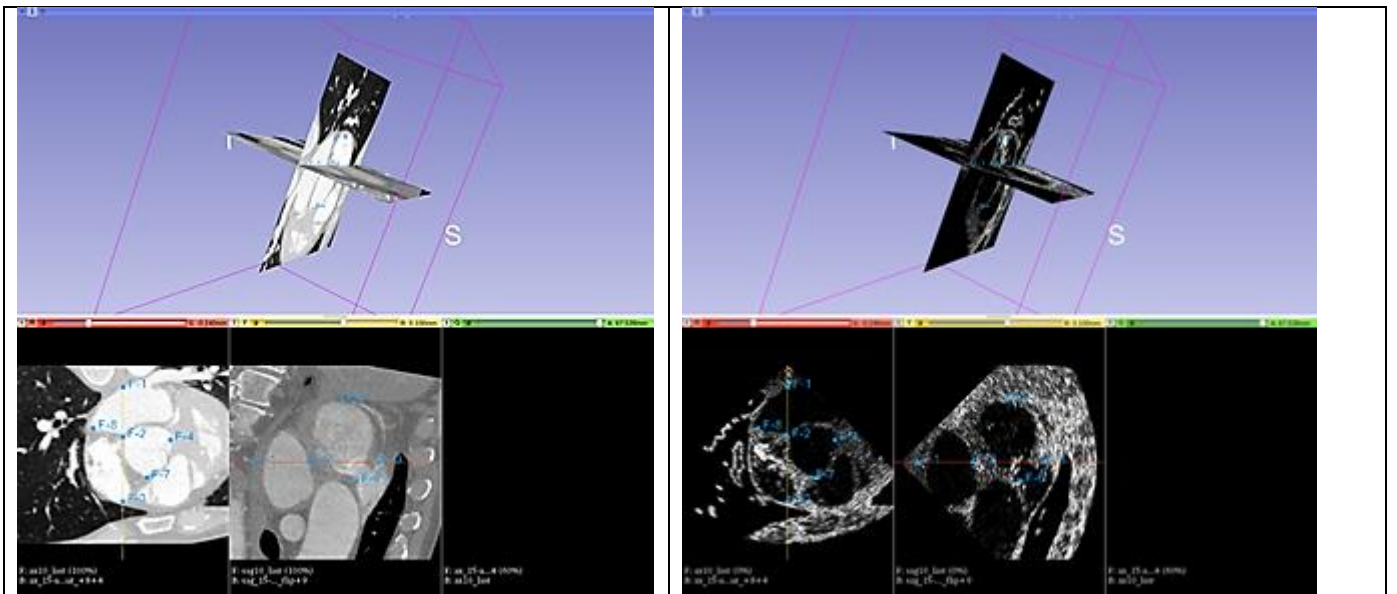


Figure 2.9 – result of the registration between the 2 perpendicular US planes (top right) and the 3D CT (top left – the 2 cut planes of the 3D corresponding to the same location as the US images). On the bottom the corresponding registered CT cut planes and 2D US images. The fiducial points used to estimate the TRE are highlighted on test slices.

Where  $d()$  denotes the Euclidian distance. The mean of the 8 TREs (mTRE) for one registration can also be considered as a global accuracy index for this registration.

On the boxplots of the 55 mTREs, Figure 2.10, we can observe that the distribution of the errors is smaller when using two perpendicular planes than just using one ( $p < 0.029$ ). As a consequence, the median error of all the mTRE was reduced from 2.54 to 1.7 mm using the two planes .

Regarding the registration accuracy, the global median target registration error (mTRE) of 1.7 mm is of the same order of magnitude as those reported in the literature: less than 5 mm for [81], 1.5 – 4.2 mm for [41] and 5.6 mm for [7].

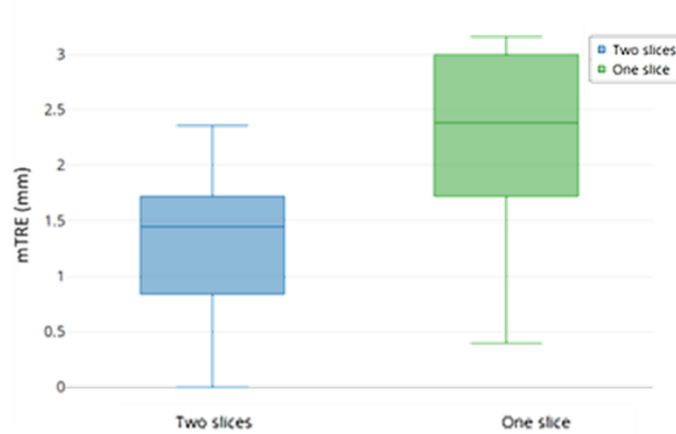
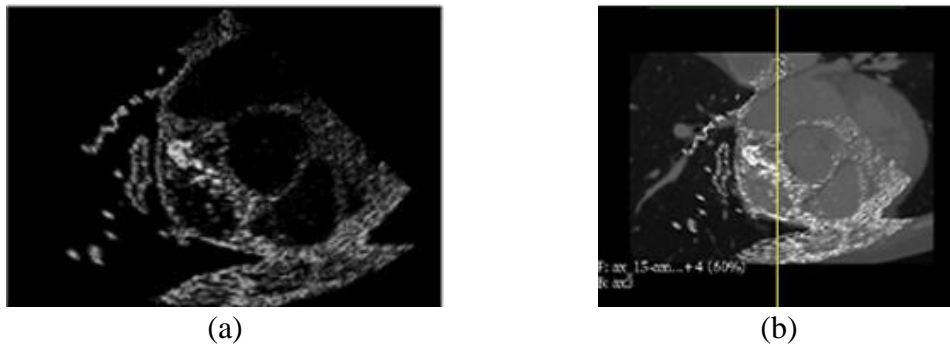


Figure 2.10 – Box plots of the mean Target Registration Error (mTRE).

### 2.4.5.3. Visual validation

Figure 2.11 shows the two perpendicular simulated US images (a and c). The estimated corresponding reformatted CT slice with the US images superimposed on them are on (b and d). The visual examination of these two figures shows a good alignment with an initial point around the ground truth (GT), Some higher accuracy could probably be gained by considering the estimated pose as a starting point closer to the ground truth.



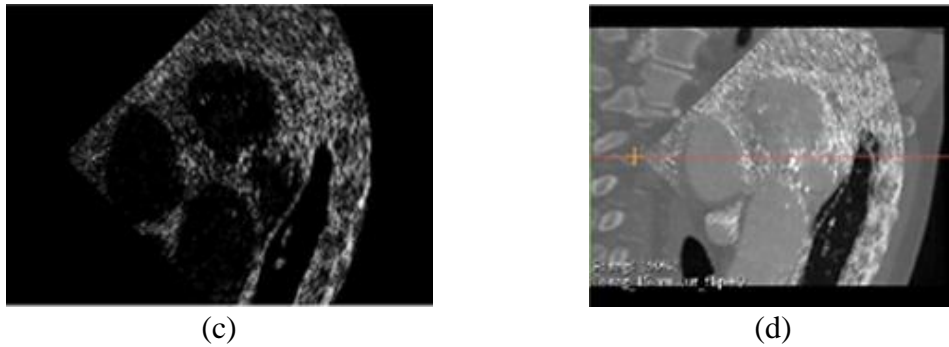


Figure 2.11 – Visual validation On (a) and (c), the 2 perpendicular simulated US images to be registered. On (b) and (d) the corresponding CT planes estimated by the registration. The US images are superimposed on these CT slices.

## 2.5. Discussion

From the above quantitative results, we can conclude that, on the one hand, the registration accuracy obtained using two perpendicular US images gave us a better result in terms of TRE compared to one US image (median TRE decreasing from 2.54 to 1.7 mm). The global median target registration error (TRE) of 1.7 mm is also on the same range of magnitude as those reported in the literature.

These results were obtained from simulated US images. We are fully aware that there are differences between simulated and real US images (signal attenuation compensation, acoustic shadowing, post processing of real US images...). However, we found in the case of the study conducted before our study [7], that a method developed on simulated data performed also well on real data and we are therefore confident that our method will also work on real data.

This study demonstrates the interest of using two perpendicular US planes, in terms of a more accurate pose localization within the CT. This will allow the radiologist to have more precise control of the therapy.

The registration was performed offline after all the US data had been acquired (or simulated). The mean intraoperative registration time was approximately 6 seconds using just one US image and 7.5 seconds using the two perpendicular US images. This computation time is not suitable for a real-time application, for this reason, we are going to present in the next chapters some convolutional neural networks (CNNs) registration framework for transesophageal ultrasound/computed tomography image registration to solve the problem of high computation time of the classical iterative methods.

## 2.6. Conclusion

In this chapter, the state-of-the-art for slice to volume registration using the classical iterative methods and their applications in medical failed has been presented. In this context, image processing methods have been proposed to improve the planning and the guidance of the therapy. We proposed a two perpendicular 2D CT/3D CT registration approach adapted to the guidance of the transesophageal HIFU therapy. We performed rigid registration of two 2D planar echocardiography images within a cardiac 3D CT volume. The results indicated a promising accuracy of the proposed technique.



## Chapitre 3

# Learning-Based Registration: supervised Transformation Estimation

### 3.1. Introduction

Image guided interventions are one of the most important applications for image registration, which helps doctors to save many patients' lives. On the other hand, this problem is considered as one of the most complex and complicated issues to tackle in medical image processing since it needs accuracy and speed at the same time.

As we already described in chapter 2, image registration is fundamental to the image-guided intervention *e.g.*, telesurgery, image-guided radiotherapy, HIFU image guidance therapy, ... because most of them cannot be operational without using image registration techniques [82]. For example, in an image guidance therapy the treatment planning is established on diagnostic or pre-interventional images (typically high-quality 3D image), on which the treatment planning is conducted, needs to be registered on an intra-operational image ( 2D, low-quality and noisy in the case of ultrasound) so that the procedure can be performed with maximum precision and minimum risk of irradiation of adjacent healthy organs in image-guided radiotherapy, as well as to preserve the surrounding tissue in a minimally-invasive ablation. In this type of procedure there are different challenges when trying to merge the different types of information: the modalities can carry totally different information (2D/3D, X-ray attenuation coefficient/echogeneity, ...), the intra-operative modalities are generally of poor quality and/or noisy, the morphology of the organs can change between the two modalities (deformations due to the gesture, to the respiration, to the cardiac beats, ...). These problems can largely compromise the quality of the image guidance. In practice, these different problems must be taken into account and other image processing techniques must be associated with the registration, which makes the problem very difficult and complicated [83].

In addition to these problems that can impact the feasibility and quality of registration, the calculation time is also one of the big issues in interventional image-based guidance. Iterative registration methods as described in chapter 2 are very time consuming, which is not suitable for real time image registration. Recently, huge advances in the field of machine learning and deep learning have enabled the implementation of deep neural networks in medical applications, where image registration has been the focus of new work that has accelerated and increased the performance of registration over traditional iterative intensity-based techniques.

Based on the literature a taxonomy of deep-learning-based registration methods into two categories can be proposed: deep similarity metrics and deep learning for transformation estimation.

These two categories of approaches can further be divided into supervised, weakly-supervised and unsupervised approaches, based on the learning paradigm used to train the networks. The nature of the transformation to be estimated may affect directly the level of supervision of the methods. The ground truth is much easier to synthesize for a rigid/affine registration than for an elastic registration. Indeed, in the case of rigid/affine registration, the training data can be generated by random combinations of operations such as rotation, translation and scaling. Furthermore, unlike non-rigid transformations, the parameters of rigid transformations are global and can be set manually.

In conclusion, rigid registration is generally easier to perform than elastic registration. There are less parameters to estimate and learning is generally easier because the data is easier to find. Indeed, it is easier to acquire physical data to which a simple rigid transformation has been applied. Similarly, numerical phantoms are also easier to realize because they require less parameters.

Paradoxically, most of the work on deep learning registration deals with elastic registration. However, in this chapter we will more focus on supervised rigid image registration using convolutional neural network. We will start by summarizing the latest development in deep learning based medical image registration, followed by the contributions and finally our proposed framework, evaluation and results.

### **3.2. Background**

The fundamental components of image registration are identical in both traditional and deep learning DL-based approaches, namely a similarity metric, transformation model and an optimiser. Neural networks have been integrated into this framework to replace/enhance the role played by one or more of these components. We can classify deep learning image registration methods into three main classes, namely approaches that (a) use neural networks as the similarity metric (often referred to as deep similarity); (b) parameterize the transformation model using neural networks; and (c) use neural networks to facilitate other operations (such as feature extraction or learning new image representations) which improve registration quality.

In our work we propose the following process: we first run the moving and fixed input image pairs through a Siamese architecture composed of convolutional layers, thereby extracting the features from the moving and fixed maps analogous to dense local descriptors (use neural networks to facilitate other operations), then match the feature maps, and finally run these joint feature maps through a registration network, which directly outputs the set of the rigid registration parameters set (parameterize the transformation model using neural networks). So, in this chapter we will focus on using networks to assess the rigid transformation parameters.

Based on the literature, five kinds of deep neural networks have already been applied to medical image registration, namely convolutional neural network CNN, Staked

Auto-Encoders (SAEs), Recurrent Neural Network (RNN), Deep Reinforcement Learning (DRL) and Generative Adversarial Network (GAN).

### 3.2.1. Convolutional neural network

CNNs should be considered as the bases of deep learning techniques, in which the whole given image (or some extracted patches) is fed directly to the network. Contrary to classical neural network-based image processing approaches whose goal is to extract only certain features from the image, the CNN-based registration approach tries to detect pairs of structural features on both fixed and moving images and attempt to align them. The detection and selection of these structural features are fully automated using CNN.

As shown in Figure 3.1 a typical CNN has some interleaving kernel and pooling layers and is ended by a typical two- or three-layer fully connected network. The kernels are trained to extract the most significant features by convolution with the input, while the pooling layers decrease the course of dimensionality, and make the results invariant to the different geometrical transformations. The output of each layer, so-called a feature-map, is passed to the next layer. When the number layers are high, a hierarchical feature set can be obtained, and the network can be considered as a deep CNN. The feature-maps from the last layers are concatenated and vectorized to feed a fully connected two or three-layer network for the final result.

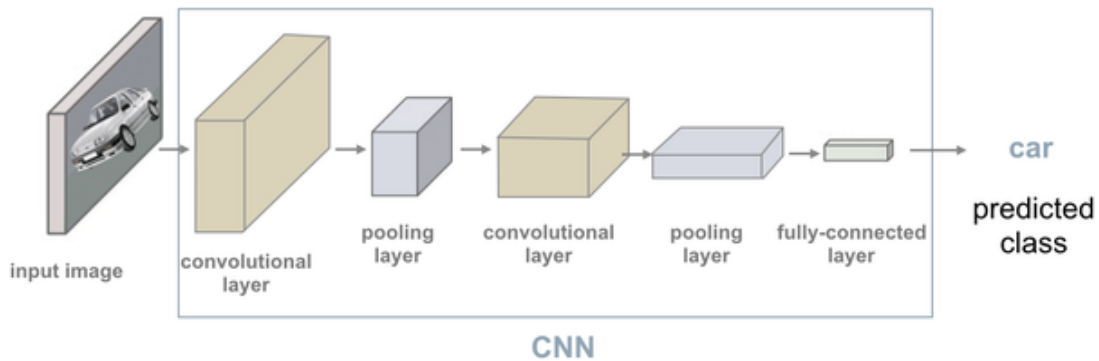


Figure 3.1 – Example of Convolutional Neural Network (CNN) architecture.

Several CNN architectures have been proposed in recent years, each with specific architectural modifications to address the issue of vanishing/exploding gradients common to deep networks, such as AlexNet [84], VGG [85], ResNet [86], and DenseNet [87]. Among these, in the field of medical image segmentation and registration, the most widely used architecture is the U-Net [88], which we will discuss in more detail in the next chapter.

### 3.2.2. Autoencoder

An autoencoder (AE) is a kind of neural network that learns to copy its input to its output without supervision [89]. An AE typically consists of an encoder that encodes the input into a low-dimensional latent state space and a decoder that restores the original input from the low-dimensional latent space. To prevent an AE from learning an identity function, regularized AEs have been invented. Examples of regularized AEs include the sparse AE, the denoising AE and the contractive AE [90]. Recently,

convolutional AE (CAE) has been proposed to combine CNN with traditional AEs [53]. CAE replaces the fully connected layer in the traditional AE by convolutional and transpose convolutional layers. CAE has been used in many medical image processing tasks such as lesion detection, segmentation, image restoration [91]. In contrast to the above mentioned AEs, variational AE (VAE) is a generative model that learns the latent representation using a variational approach [92]. VAE has been used for anomaly detection [93] and image generation [94].

### **3.2.3. Recurrent neural network**

A recurrent neural network (RNN) is a kind of neural network that is used to model dynamic temporal behavior [95]. RNN is widely used for natural language processing [96]. Unlike feedforward networks such as CNN, RNN is suitable for processing temporal signals. The internal state of RNN is used to model and ‘memorize’ previously processed information. Therefore, the output of RNN depends not only on its immediate input but also on its input history. Long short-term memory (LSTM) is a specific type of RNN that is used in image processing tasks. Recently, Cho et al proposed a simplified version of LSTM, called gated recurrent unit [97].

### **3.2.4. Reinforcement learning**

Reinforcement learning (RL) is a type of machine learning that focuses on predicting the best actions to take based on the current state in an environment [93]. RL is usually modelled as a Markov decision process using a set of environmental states and actions. An artificial agent is trained to maximize its cumulative expected rewards. The training process often involves an exploration-exploitation trade-off. Exploration involves exploring the whole space to gather more information while exploitation involves exploring the promising areas given the current information. Q-learning is a model-free RL algorithm, which aims to learn a Q function that models the action-reward relationship. The Bellman equation is often used in Qlearning for reward calculation. The Bellman equation calculates the maximum future reward as the immediate reward the agent gets from entering the current state plus a weighted maximum future reward for the next state. For image processing, the Q function is often modelled as a CNN, which could encode the input images as states and learn the Q function via supervised training [98], [99].

### **3.2.5. Generative adversarial network GAN**

The Generative Adversarial Network (GAN) has been proposed by Goodfellow et al. In 2014 [100]. It is composed of two competing subnetworks, the generator, and the discriminator. The generator is trained on a ground-truth dataset to synthesize fake samples, while the discriminator should discriminate between fake (synthesized) data and the real one and give its result as a binary output. Based on the survival competition between the generator and the discriminator, just like the game theory, the network can be trained on a small set of data so that the generated samples cannot be discriminated anymore, and the network goes towards equilibrium. The network gets the name GAN because the generator is trained in an adversarial manner based on the feedback from the discriminator. While the original GAN was applied to noise suppression in images,



it has gained increasing popularity in recent years, and has been applied to almost all medical imaging problems. In the context of image registration, the generator takes the fixed and moving images as input and tries to produce the transformation parameters such that the transformed moving image, called warped image, is indistinguishable from the ground-truth image by the discriminators. In this context Lu et al [101], proposed a method, to reduce the deformation between two 2D images, as shown in Figure 3.2., They introduced the Cycle Generative Adversarial Network (CycleGAN) into their method simulating TEE-like images from CT images to reduce their appearance discrepancy. Then, they perform gridless registration to align TEE-like images to the real TEE ones. Experimental results on CT and EEG images of children and adults show that the proposed method outperforms other compared methods.

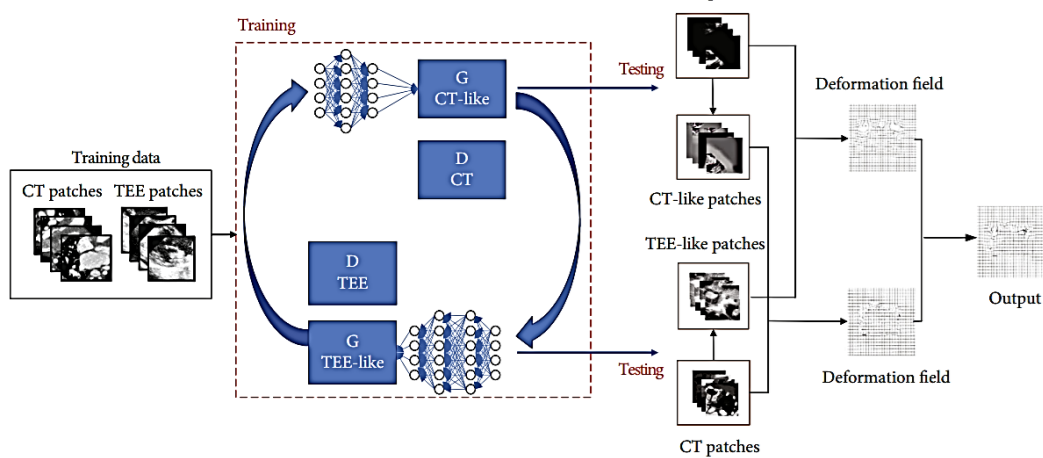


Figure 3.2 – GAN network example in image registration [102].

Also, in medical imaging, GAN has already been used to perform inter- or intra-modal image synthesis, such as MR to synthetic CT [103], CT to synthetic MR [103], [104], CBCT to synthetic CT and so on.

GAN is typically used to provide additional regularization or to convert a multimodal registration to a unimodal one. Besides medical imaging, GAN has been widely used in many other fields including science, art, games and so on.

### 3.3. Related work in learning based rigid medical image registration

In this section we will present the related work that has used CNN first, as a similarity metric, second to estimate rigid transformation parameters. Finally, we will present the latest papers on slice-to-volume learning-based registration approaches (supervised and unsupervised).

#### 3.3.1. Deep Similarity based Registration

Cheng et al [105] propose a deep similarity learning network to train a binary classifier. The network is trained to learn the correspondence of two image patches from a pair of CT-MR images. The continuous probabilistic value is used as the similarity score. Similarly, a similarity metric based on a regression CNN was proposed by Haskins et al [106] to register Magnetic Resonance Imaging (MRI) and Transrectal Ultrasound

(TRUS) images, which demonstrated promising performance compared with MI, and several other conventional similarity metrics.

In addition, Sedghi et al. [107] perform the rigid registration of 3D US/MR (modalities with even greater difference in appearance than MR/CT abdominal scans) using a 5-layer neural network to learn a similarity metric that is then optimized by Powell's method. This approach performs better than registration based on MI optimization.

### 3.3.2. Supervised transformation estimation

The motivation of transformation estimation using deep learning approach is to develop a network that could estimate the transformation that corresponds to the optimal similarity in one step.

A few approaches have focused on rigid registration of multimodal images, *e.g.*, Chee et al. [108] use a CNN to predict the transformation parameters between 3D brain MRI volumes. In their framework called Affine Image Registration network (AIRNet), the Mean Square Error (MSE) between the predicted and ground truth affine transforms is used to train the network.

In addition, Yao et al [109] use a regression CNN for a coarse 3D/3D rigid registration, which then serves as an initialization of a conventional intensity-based registration method for fine-grained registration. This approach combines so CNNs with conventional methods to align 3D CT and CBCT images.

Several papers have also explored the registration of MRI and TRUS images on prostate images. Some of the works are based on the use of two publicly available datasets RESECT [103] and BITE for this registration task. However, most of the studies on MRI and TRUS images registration are learned on private datasets. Guo et al [110] propose a supervised network to tackle rigid MRI-TRUS registration on prostate images. They propose a new strategy to generate augmented datasets and design a coarse-to-fine multistage network, which significantly reduces the registration error compared to previous methods.

### 3.3.3. 2D/3D image registration using CNN

In most of the multi-modal registration applications discussed so far, the dimension of the fixed and moving images are identical. Publicly available datasets provide 3D image volumes, which can also be employed for slice-wise 2D/2D registration. Therefore, the proposed studies so far have mainly focused on 2D/2D or 3D/3D image registration. However, 2D/3D image registration is still of essential interest for a variety of clinical applications. Thus, in contrast to the classical registration research, multimodal 2D/3D registration by DL should also be an object of study. However, this task is even more challenging, due to the difference in dimensionality and the overlapping tissues and low contrast issues common to 2D medical images such as x-rays radiographs.

Until now, studies on 2D–3D registration are mainly focused on registering to a specific 3D modality image either a protectional images (*e.g.*, fluoroscopy/MRI) or a cross section slice (*e.g.*, US/MRI).

### 3.3.3.1. Protectional images to volume image registration using CNN

Miao et al. [53] were the first to use deep learning to predict rigid transformation parameters. They used a CNN to predict the transformation matrix associated with the rigid registration of 2D/3D X-ray attenuation maps and 2D X-ray images. Hierarchical regression is proposed in which the 6 transformation parameters are partitioned into 3 groups. Ground truth data was synthesized in this approach by transforming aligned data. This is the case for the next three approaches that are described as well. This approach outperformed the classical image and optimization-based registration approaches in terms of both accuracy and computational efficiency. The improved computational efficiency is due to the use of a forward pass through a neural network instead of an optimization algorithm to perform the registration.

Liao et al [111] proposed a novel learning-based Multiview 2D–3D rigid registration method that directly measured the 3D misalignment using a Point-Of-Interest Network for Tracking (POINT) and found the point-to-point correspondence between two images.

### 3.3.3.2. Cross section slice to volume image registration using CNN

The slice to volume registration is more challenging because the 2D image contains less information of the volume (only a cross section of the volume) than in the previous case (projection of the volume on a plane). So very few papers deal with slice-to-volume registration. Among them, Salehi *et al.* [112] propose an 18-layer residual CNN regression model for 3D pose estimation, and rigidly register reconstructed fetal brain MRI images to a standard space (atlas). Then, based on images generated by the four transformations (*i.e.*, scaling, horizontal or vertical shift and rotation), they validate the effectiveness of their geodesic loss term and show the superiority of their method over the NCC-optimization-based registration.

Recently, Guo et al. [113], propose an end-to-end unsupervised frame-to-volume registration network called FVR-Net. This network is trained to register intra-operative 2D transrectal ultrasound (TRUS) with pre-operative 3D magnetic resonance (MR) volume to guide the prostate biopsy, this without requiring hardware tracking. Their results demonstrate superior efficiency of the proposed method for real-time interventional guidance with a run time of approximately 0.7ms and with a very competitive registration accuracy (the distance error being 2.73mm).

Finally, Fu et al. [114], performed the registration of 3D MRI to 2D MRI slice that was extracted from the 3D MRI after random rotation and translation. they propose an intentional overfitting deep learning-based network (ION) to perform volume-to-slice registration for MRI abdominal images.

### 3.4. Ultrasound to CT 2D Rigid Image Registration using CNN

Before presenting the registration method that we have proposed, we will remind you of the different hypotheses that allowed us to consider the registration as rigid. The heart is a moving organ. However, some characteristics of cardiac motion allowed us to consider a rigid registration scheme. First, we had at our disposal a Cine CT from a patient's thorax composed of 20 volumes at each 5% phase of the RR interval. Second, we were interested in ventricular fibrillation. During its diastolic phase, the ventricle is relatively stationary. The HIFU treatment will be fired in this phase to have a fixed focal point with respect to the organ and thus avoid a dispersion of heat prejudicial to the necrosis of the tissues. Thus, a quasi-static ventricle pose can be considered. Moreover, on our US system, the acquisition is synchronized with the ECG. Thus, it is relatively easy to create pairs of US/CT images at the same phase and so to consider rigid registration.

#### 3.4.1. Materials and method

In this section, we will introduce our proposed framework for estimating the transformation parameters of a rigid image registration between a preoperative CT slice and an intraoperative US image. In our case and following the approach described in [115], only a 2D rigid transform with three Degrees of Freedom (DOF) – one rotation and 2 translations- was to be estimated.

The main idea is to estimate the registration that best aligns some common characteristics or features of the images. The information contained in the two images is very different in nature (gray levels proportional to the X-ray absorption coefficient of the tissues for CT and information formed by the reflection of ultrasonic waves on surfaces and speckle for US). It is therefore necessary to extract from both imaging modalities some common information (in our case the shapes of the organs) before performing the registration. We therefore proposed the following registration framework (Figure 3.3):

- (i) Descriptors are initially extracted from both the moving  $I_M$  and the fixed  $I_F$  images using Deep Learning. For this,  $I_M$  and  $I_F$  are passed through a Siamese CNN architecture consisting of some convolutional layers, thus extracting two feature maps  $f_F$ , and  $f_M$  which are analogous to dense local descriptors.
- (ii) These feature maps are then combined together in a concatenating layer.
- (iii) This image of the corresponding concatenated feature maps is fed as an input into a convolutional registration network which directly outputs the rigid registration parameters set T (two translations, one rotation) of the rigid registration.

This framework should be trainable end-to-end for the rigid registration task.

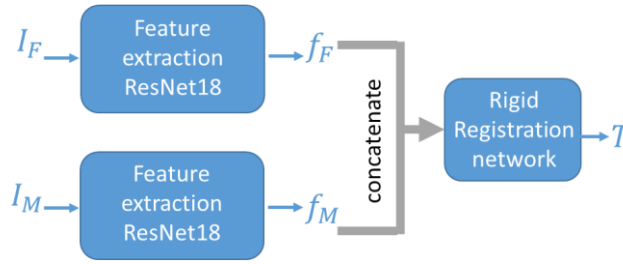


Figure 3.3 – The overall of the proposed framework.

### 3.4.1.1. Feature extraction

The first step of the framework is feature extraction. Features extraction is a classical tool in deep learning. One common feature extraction technique is to feed the image to a conventional pre-trained neural network and use the representation for that particular image in the intermediate layers of the neural network. We used the ResNet18, which is one of the most efficient standard feature extraction models that can be used in many medical application [116]. The advantage of this model is that it handles the vanishing or exploding gradient problem when the CNN goes deeper.

Resnet18 can be found implemented in PyTorch. This implementation offers a version with the weights pre-trained for feature extraction on ImageNet, the large benchmark database.

In our case, each of our input modalities has its own image characteristics. Thus, we passed each of the two images to be registered in its own network but in a Siamese manner. In a Siamese network, both models are instances of the same model (same weights and structure). The Siamese network allows to integrate the classification problem and the similarity problem. The network is trained to minimize the distance between samples of the same class and to increase the distance between classes. There are several types of similarity functions through which the Siamese network can be trained. In our case we used the L2 loss function.

### 3.4.1.2. Matching

These two feature maps needs be combined across images as a single tensor to feed it into the rigid transformation parameters estimation network. To achieve this, a concatenation of descriptors along the channel dimensions is performed in a concatenation layer.

### 3.4.1.3. Registration network

We will present the registration network architecture which consists of three blocks of convolutional layers using a kernel size of 5, each followed by batch normalization layers, and a rectified linear unit (ReLU). The last layer is a fully connected layer for estimating the rigid registration parameters as shown in Figure 3.4. The network receives as input the concatenated map of the extracted features from moving and fixed images, and directly estimates the parameters ( $t_x$ ,  $t_y$ , and  $\vartheta$ ) of the rigid transformation that links these feature maps. The idea behind this architecture is that the estimation is performed in a bottom-up manner where the early convolutional layers vote for candidate transformations, and these are then processed by the later layers to aggregate the votes.

The first convolutional layers can also enforce local neighborhood consensus by learning filters that only fire if nearby descriptors in image  $I_M$  are matched to nearby descriptors in image  $I_F$ .

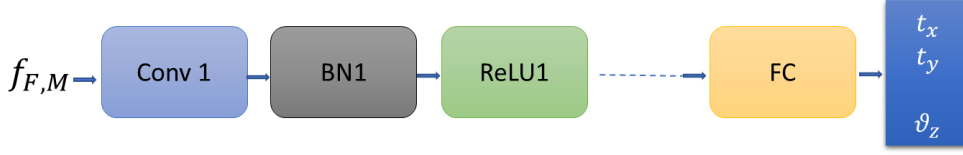


Figure 3.4 – Architecture of the regression network.

#### 3.4.1.4. Training

We considered a supervised learning scheme. We had at our disposal a learning dataset composed of corresponding pairs of US and scanner images whose geometric transformation was known (the ground truth GT). This allowed us to simply formulate the learning loss function as the L2 norm of the error between the GT ( $T_{GT}$ ) and the predicted transformation parameter ( $T_{Est}$ ).

$$L = \alpha \|t_{GT} - t_{Est}\|^2 + \beta \|\vartheta_{GT} - \vartheta_{Est}\|^2 \quad (3.1)$$

With  $t_{GT}$  and  $t_{EST}$  are the translation vector of respectively the ground truth transformation and the estimated one expressed in mm:  $\vartheta_{GT}$  and  $\vartheta_{EST}$  degree; and  $\alpha$  and  $\beta$  are weights controlling the balance between the translation and the rotation losses. The choice to use mm for translation and degrees for rotation allowed to ensure a certain coherence and normalization between the translation and rotation parameters. Indeed, an error of 1 degree in rotation leads to a displacement of 1 mm at 60 mm from the center of rotation, *i.e.*, half the depth of the ultrasound beam. Because of this relative coherence between parameters,  $\alpha$  and  $\beta$  could be fixed at 1

For training the network, we computed the gradient of the loss function with respect to the estimated rigid parameters ( $t_x$ ,  $t_y$ , and  $\vartheta$ ). This gradient is then used to minimize the loss function by using backpropagation and Stochastic Gradient Descent.

After training, the network can be applied for registration of unseen image pairs. We implemented the network using PyTorch and we trained it on a NVIDIA TitanX GPU with 10000 iterations, and batch size of 16, which took approximately 12 hours.

#### 3.4.2. Datasets

Training data is one of the key points of any learning-based method. Medical image registration problems are usually quite complicated to learn because the transformation (the Ground Truth, GT) between images to be registered is rarely known on real images. This hinders supervised learning methods.

To compensate for this lack of ground truth, we decided to simulate US images from CT data to which a formally known transformation can be applied. In conclusion, our study has been conducted using real CT datasets and corresponding simulated US images.

### 3.4.2.1. CT datasets

Our study has been conducted on MM-WHS 2017, a public available dataset for multi-modality whole heart segmentation [5], [6]. Figure 3.5, Figure 3.6 show examples from some CT volumes of the datasets.

All the data were obtained from two state-of-the-art 64-slice CT scanners (Philips Medical Systems, Netherlands) using a standard coronary CT angiography protocol at two sites affiliated with Shanghai Shuguang Hospital. The volumes were acquired in the axial view, covering the whole heart from the upper abdomen to the aortic arch. The in-plane resolution was about  $0.44 \times 0.44$  mm, and the average slice thickness was 0.60 mm. In these volumes, the esophagus was manually coarsely segmented.

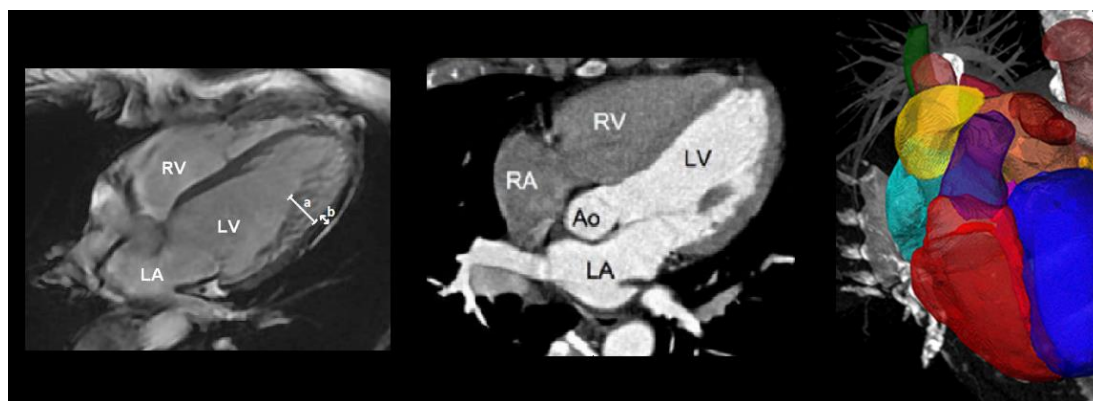


Figure 3.5 – Volumes from MMWHS2017 datasets.

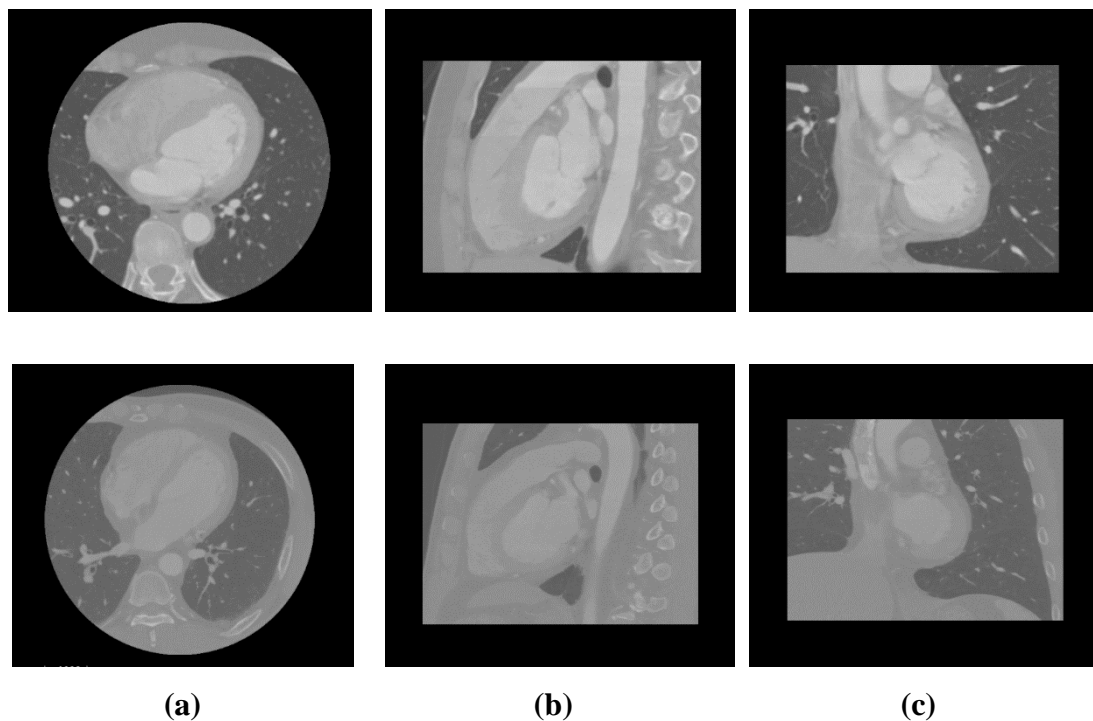


Figure 3.6 – Thorax from the superior vena cava to the stomach. Axial (a), Sagittal (b) and Coronal (c) views extracted from two CT volumes.

### 3.4.3. US datasets

From the MMW- HS2017 dataset CT volumes, we create a set of corresponding pairs of CT and US images with known transformation (Figure 3.7)

- 1) Extraction of the CT 2D slices.

First, we extracted randomly 4000 2D CT oblique cut planes from the 20 CT volumes (200 image per volume). For this we choose randomly 4000 initial poses along the esophagus axes within the 20 CT volumes. For each pose, we create a new referential by setting some randomly transformation near these initial poses with some translations within  $\pm 10$  mm and rotations within  $\pm 15$  degree around each coordinate axis. The  $x$ - $y$  plane of this referential will serve as the fixed 2D CT image  $I_{ct}$  (Figure 3.8.a). The origin of the  $x$ - $y$  plane served also as origin of  $I_{ct}$ .

- 2) Definition of the ground truth transformation.

For each 2D CT slice, we randomly define the pose of the simulated US probe origin by setting some randomly defined 2D translation within a range of  $\pm 10$  mm and one rotation in a range of  $\pm 15$  degree from the origin of  $I_{ct}$  (Figure 3.8.a). These two translations and one rotation define the ground truth rigid 2D transformation  $T_{GT}$  between the 2D CT slice and the US image.

- 3) Simulation of the 2D US image.

Once we have defined the pose of the simulated US probe origin, we simulated the corresponding US image with the method described in [3] that predict the appearance and properties of a B-scan ultrasound image from a probe origin pose, the point spread function of the US device, the acoustical impedance of the tissues and some tissue-adapted distribution of point scatterers which gave the speckle (Figure 3.8.b). In our case we mimicked a 128 elements TEE probe working at a frequency of 7 MHz, with a beam angle of 90 degree and a depth of 150 mm.

- 4) After the simulation, every CT and US image pair is resampled to the same spatial resolution in mm according to the spacing information. Finally, the gray intensities of both images were scaled between  $[0,1]$ . At the end, for each of the 4000  $I_{CT}$  we get an US image  $I_{US}$  and a transformation ground truth  $T_{GT}$

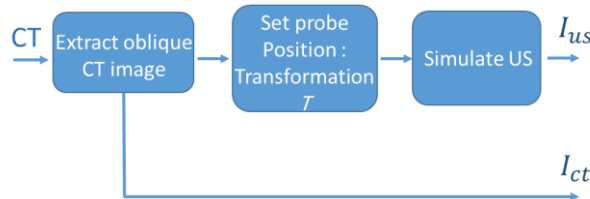


Figure 3.7 – Dataset's creation workflow.



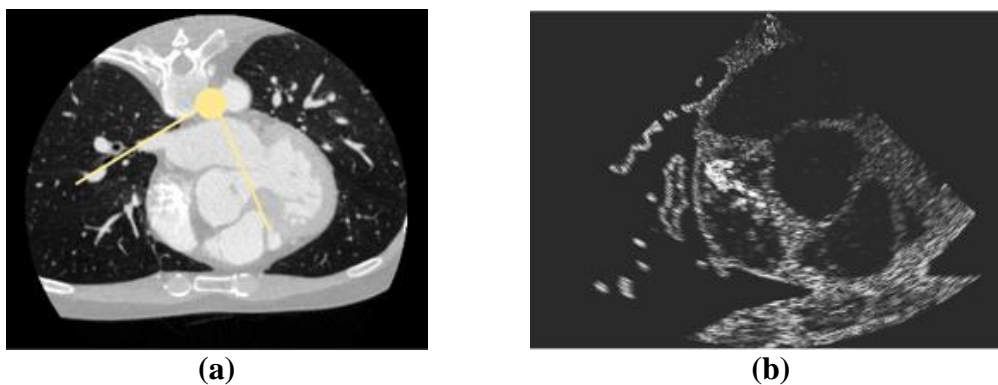


Figure 3.8 – Simulation of a CT/US image pair: a) 2D CT slice extracted from a volume with the pose and field of view (yellow) of the US probe; b) the simulated US image.

From this dataset, the network was trained by selecting the 3600 pairs of corresponding  $I_{ct}$  and  $I_{us}$  slices from 18 of the 20 cardiac CT scans. For validation, we used 400 image pairs from the 2 remaining volumes.

Figure 3.9, and Figure 3.10 show some examples of the extracted CT slices and the corresponding simulated US images slices.

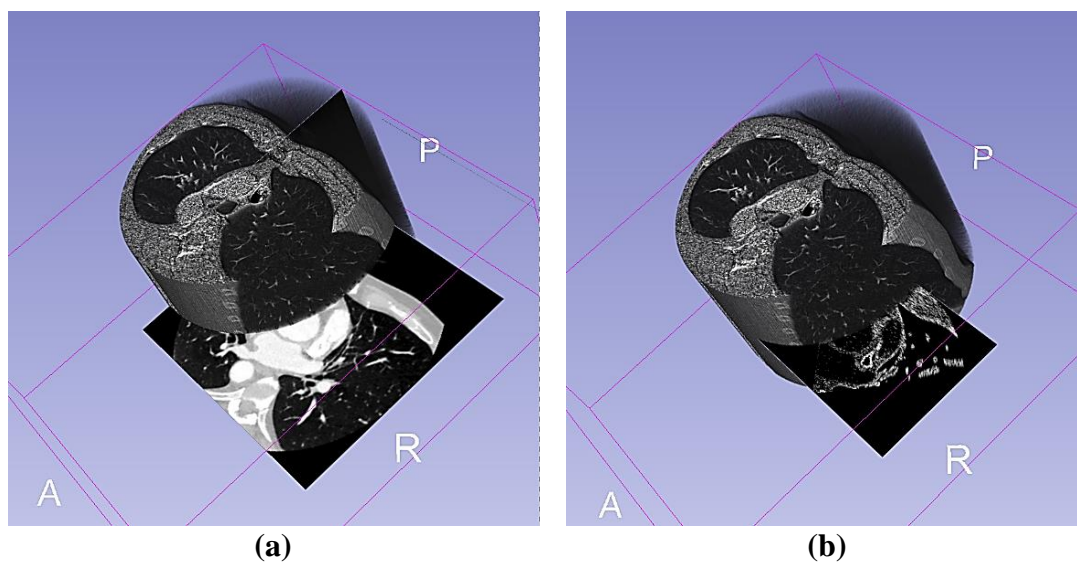


Figure 3.9 – An example of the extracted CT slices and the corresponding simulated US images in the 3D CT volume.



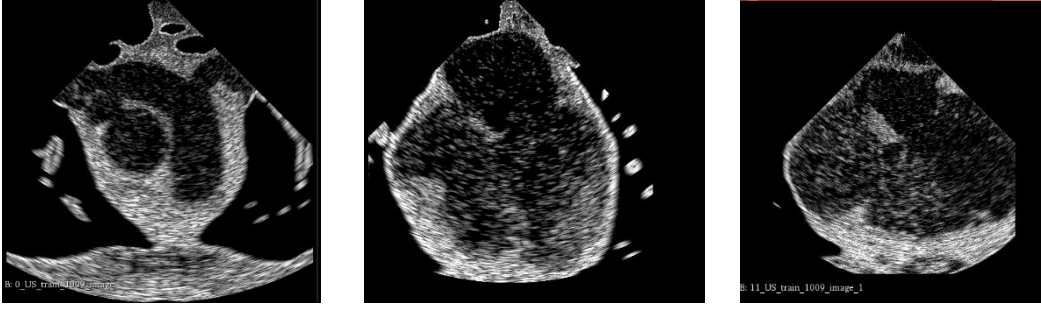


Figure 3.10 – Three examples of the extracted CT slices and the corresponding simulated US images slices.

### 3.4.4. Evaluation

We compared the registration results obtained by the proposed methods to those obtained by the classical iterative rigid registration method implemented in the SimpleITK Library [77]. As a reminder, the iterative method estimates by optimizing the parameters of rigid transformations that tries to maximize a similarity measure between the moving and the fixed image. In our study, we used Normalized Mutual Information (NMI) as a similarity measure for the iterative method. Indeed, during a comparative study between similarity metrics, this measure proved to be one of the most suitable for our CT/US registration problem [4]. The result of this iterative method is also the set of the 3 transformation parameters (2 translations and one rotation angle).

#### 3.4.4.1. Computation time

For our deep learning-based method, the average registration computation time for all the 400 image pairs is now less than 3 ms for each image pair. This low computation time allows us to consider a real-time application.

For comparison, the classical iterative method takes about 6 seconds to register a pair of images.

#### 3.4.4.2. Transformation estimation error

We compared the parameters of the transformation obtained by our proposed methods with those of the ground truth (GT). A transformation is composed by a translation vector  $t$  ( $t_x, t_y$ ) and a rotation ( $\vartheta_z$ ). We separately evaluate the translation errors and the rotation errors between the estimated pose of each of the 400 validation image pairs and their associated GT.

The translation error is measured by equation (3.2), where  $t_{GT,i}$  and  $t_{Est,i}$  are the translation vectors of respectively the GT and the estimated one.

The rotation error is given by the angle difference in degrees between the GT pose orientation angle and the estimated rotation angle (equation (3.3)), where  $\vartheta_{GT,i}$  and  $\vartheta_{EST,i}$  are the angles that encode the orientation parameter of the pose of respectively the GT and the estimated rotation.

$$t_{Error} = \|t_{GT,i} - t_{EST,i}\|^2 \quad (3.2) \quad \vartheta_{Error} = \|\vartheta_{GT,i} - \vartheta_{EST,i}\|^2 \quad (3.3)$$

Figure 3.11.a shows the boxplots of the 400 translation errors from our CNN-based registration and the classical iterative one. The median translation errors are 1.1 mm using CNN and 1.2 mm using the classical approach.

The boxplots of the 400 rotation errors in Figure 3.11.b show that the median rotation errors are 2.1 degree using CNN and 2.4 degree when using the iterative classical method.

The Wilcoxon statistical test estimated by transformation estimation pair of the classical and CNN approaches was applied to show that the groups are significantly different.

As expected, that the results of the estimated translation parameters of the classical and CNN approaches are not significantly different ( $p > 0.43$ ).

We note that the results of the estimated rotation parameters of the classical and CNN approaches are not significantly different ( $p > 0.29$ ).

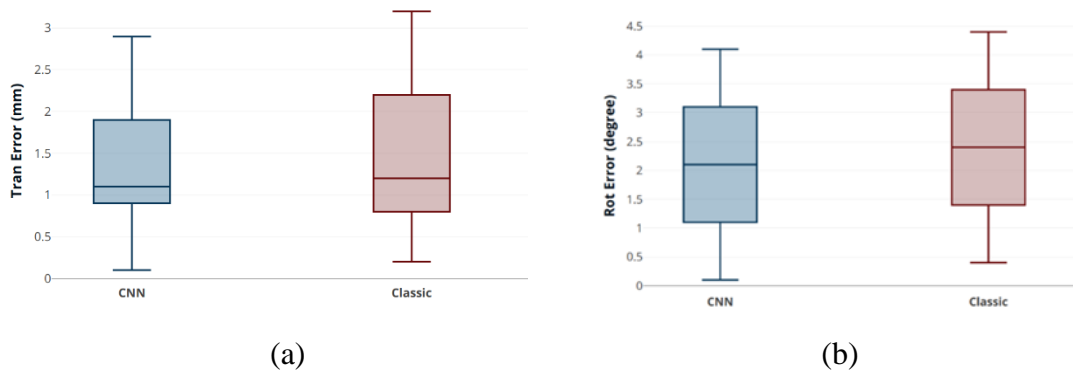


Figure 3.11 – Box plots of a) the Translation Estimation Errors, b) the Rotation Estimation Errors.

### 3.4.4.3. Target Registration Error (TRE)

Evaluation of registration accuracy can also be done by estimating the registration errors on some fiducial markers. To quantify the error, we defined eight specific feature points (or landmarks)  $P_j$  in the US fixed images as presented in Figure 3.12.

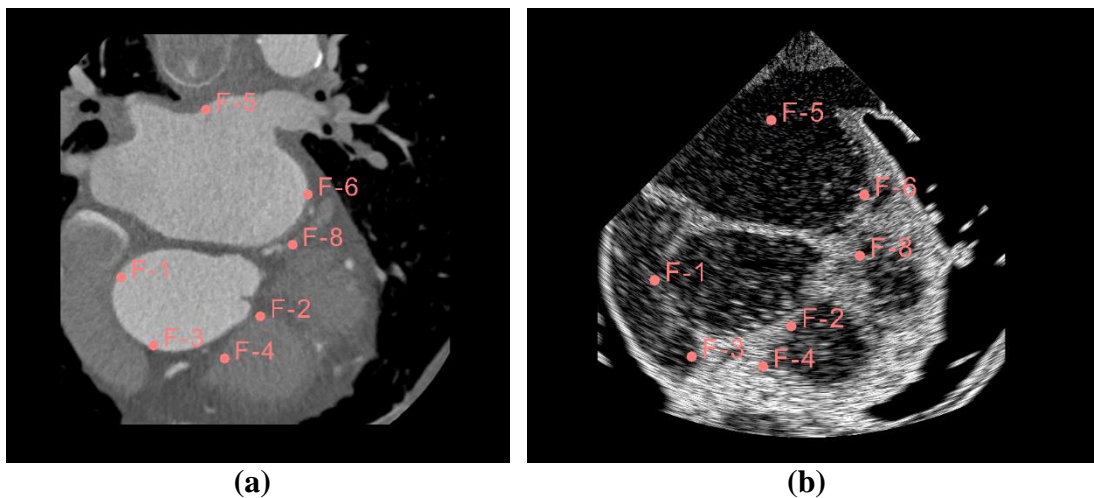


Figure 3.12 – the position of the fiducial points marked in pink.

Then we used the two transformations matrices, the estimated  $T_{Est}$  and the GT  $T_{GT}$  one, to project these points into the corresponding CT images:  $P_{Est,j} = T_{Est}P_j$  and :  $P_{GT,j} = T_{GT}P_j$ . The Euclidean distance between the corresponding projected points gives the TRE:

$$TER = \|P_{GT,j} - P_{Est,j}\| \quad (3.4)$$

Figure 3.13 the boxplot of the TREs for all the 8 fiducial points of all the 400 test images. The quantitative results show a median TRE of 2.2 mm for all the fiducial points of all the 400 test images using CNN, and 2.7 mm using the classical method. We also found that the results of the estimated TRE of the classical and CNN approaches are not significantly different ( $p > 0.28$ ).

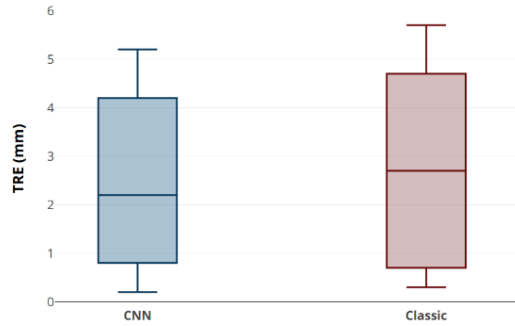


Figure 3.13 – Box plots of the Target Registration Errors.

#### 3.4.4.4. Visual validation

Figure 3.14 shows a visual comparison between a) the moving CT image and b) the simulated fixed US image pair. It shows the overlap between the moving CT image and the fixed US image: c) before registration d) after registration with the proposed method. Visually, the results obtained by the proposed method seem to provide a good alignment, this can be seen for example at the probe center, and at the bottom of the image on the thoracic chest.

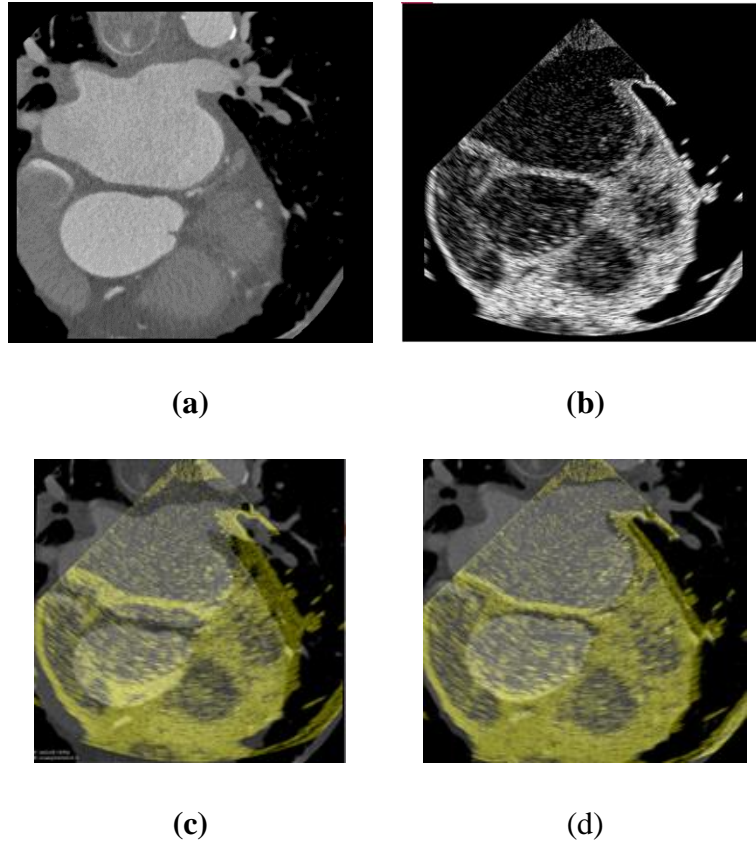


Figure 3.14 – An example of the registration of an image pair a) CT image, b) US image before registration. The overlap between the moving CT image and the fixed US image (yellow image) c) before registration, and d) after registration with the proposed method.

### 3.4.5. Discussion

In this work, we presented a deep feature learning-based approach for the registration of transesophageal US/CT cardiac images. The results showed a strong improvement in terms of computation time without degradation and even a slight improvement in terms of registration accuracy.

From the previous quantitative results, we can conclude that on the one hand, the registration accuracy obtained by CNN is of the same order as that obtained by the classical iterative method. The results obtained by CNN are even slightly better, even if statistically this improvement is not significant. Compared to other methods of the literature, the global target registration error (TRE) of 2.2 mm is on the same range of magnitude as those reported in [115]. On the other hand, the CNN greatly accelerate the processing time. The registration between two images takes only 3 ms (instead of 6 s for the classical iterative method). This gain in computation time allows us to consider implementing the 3D CT/2D US registration technique proposed by [115] in clinical practice.

These results were obtained from simulated US images. We are fully aware that there are differences between simulated and real US images (signal attenuation compensation, acoustic shadowing, post processing of real US images...). However,

we found in a previous study that a method developed on simulated data gave good result on real data [115] . We are therefore confident that our method will also work on real data.

In the next section, we will integrate the features learning approach to a minimally-invasive HIFU procedure to improve the therapy planning and guidance. We will apply our approach on a 2D/3D learning-based registration to refine the estimation of the transesophageal probe pose placement in the 3D preoperative volume.

### 3.5. Deep learning-based for slice-to-volume image registration

In this section, we will present a first attempt of an end-to-end framework to address the challenging problem slice-to-volume registration for image guidance therapy purposes. This study will be a preliminary study to bridge the gap of real time of 2D US/3D CT fusion for cardiac arrhythmia therapy.

The main goal of this study is to estimate the 3D pose of the 2D image plane in the 3D volume, therefore, to find the 3D transformation matrix that will define the pose of the transformed moving image with six dof (three translation and three rotations along x, y, z axis respectively). So, in our specific case the 2D-3D registration consists of finding the pose of the intraoperative TEE imaging plane (TEE probe position) inside the preoperative 3D CT volume, in which the ablation path has been defined.

The framework takes a 2D ( $H \times W$ ) image plane, and 3D ( $H \times W \times D$ ) volume as input. We also have at our disposal a broad estimation of an initial pose located in the central line of the esophagus. The goal is to estimate the transformation parameters  $T$  that best align these two images. From the literature we found that rigid registration with 6 degrees of freedom is suitable for 2D/3D image registration applications. Thus, the outputs  $\theta$  contains 6 degrees of freedom, *i.e.*,  $\theta = t_x ; t_y ; t_z ; \vartheta_x ; \vartheta_y ; \vartheta_z$ , including the translations and rotations along the three axes x, y, z respectively. In our case, the transformation refers to the 3D translation (in mm) and the 3D rotation (in degrees) with respect to the first initial pose

As for the 2D/2D case, we decided to separate the global framework into two sub-problems: The extraction of input image features followed by the estimation by a neural network of the parameters that best align these features after concatenation. As for the 2D/2D case we used a Resnet model as registration network to estimate the parameters.

The main challenge was to devise parallel feature extraction networks. Indeed, contrary to the 2D/2D case, now the input images do not have the same dimension, and the features must be described in a space of the same dimension in order to be concatenated and delivered to the registration network. Once this problem is solved, the framework consists again of two branches of networks to extract the features from each input, a concatenation layer to combine both images’ features”, and finally the registration network that’s will directly estimate the 6 transformation parameters as shown in Figure 3.15.

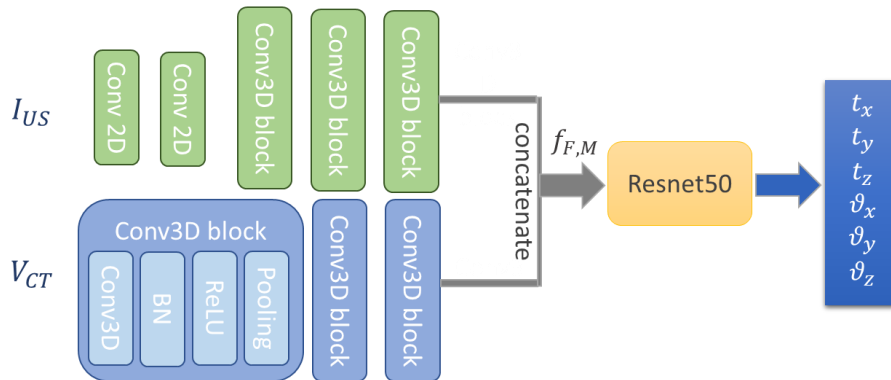


Figure 3.15 – Our proposal framework.

### 3.5.1. Feature extraction

The different dimensionality between the 2D image slice and the 3D volume is the principal challenge for the registration network performance, actually we can't directly or early concatenate the 3D moving, and the 2D fixed images, the network will totally ignore the 2D image content comparing to the volumetric information.

The features must be described in a 3D volume. For the ultrasound image the transition from 2D to 3D is done as follows:

We chose to use a two-branch feature extraction network to match the information between the two inputs, making the model sensitive to both the slice and volume information, the features must be described in a 3D volume. For the ultrasound image the transition from 2D to 3D is done as follows: The network is designed to process the 2D slice via a first 2D pathway which consists of two 2D convolutional layers to extract the low-level features from the input image plane, to extend the channel number to 3D which can then be followed by three 3D convolutional block, so that the size of the feature map matches the size of the input volume, so balancing the data information. The 3D CT volume is processed via a separate 3D pathway which consists of three 3D convolutional blocks. Each 3D convolutional block consists of a 3D convolutional layer, followed by batch normalization, a rectified linear unit (ReLU) and a max pooling layer.

We use the same hyper-parameters for each branch to maintain an end-to-end identical feature map size throughout the 2D image to the 3D volume registration framework.

### 3.5.2. Concatenation and registration network

The extracted feature maps from the 2D and 3D pathways were concatenated along the depth dimension and processed by the registration network to predict the six DOF parameters.

For the 3D plane pose estimation, we used a 50-layer residual CNN, and two fully connected layers, the last fully connected layer has size of six, which correspond to the translation and rotation parameters to estimate the slice pose in the 3D volume.

### 3.5.3. Datasets and implementation details

The proposed framework is evaluated on real 3D CT volumes, and simulated US images.

Our framework using the volumes of the MMWHS2017 datasets[5], [6], This dataset contains 60 3D CT volumes. 70% of the dataset is randomly selected for training and the rest 30% used for testing. All volumes are processed to be isotropic with mean dimensions of  $512 \times 512 \times D$  voxels.

In real- time surgery we will have some information about the endoscope tip location (inserted length, visual analysis of the image sequence during navigation, fluoroscopy, etc.) to define a candidate aera along the esophagus centerline in which the image transducer center can be located. The size of this aera is roughly 10 mm along the esophagus, so we are able to extract a sub-volume (three sub-volume per volume) of



size  $512 \times 512 \times 32$  in which the 2D US image will be. The fact to estimate the parameters in a sub-volume has several benefits: it allows to reduce the search space of the network to find the optimal transformation. It also reduces the memory load during the training phase and a strategy for data augmentation.

We used this fact to produce some Ground-Truth 2D US/3D CT sub-volumes pairs for training and evaluation. For each volume, we perform some depth sampling along the esophagus. For each of this sampled position we extract a CT sub-volume  $V_m$  we choose an initial transform  $T_{init}$ , starting from initial translation (position) in the middle of  $V_m$  of size  $512 \times 512 \times 32$ . For each sub-volume  $V_m$  we define an initial transform  $t_{init} = (t_{x_{init}}, t_{y_{init}}, t_{z_{init}})$ , and initial rotation  $R_{init} = (R_{x_{init}} = 0, R_{y_{init}} = 0, R_{z_{init}} = 0)$ . The US image is then created by simulation: 1) we define an oblique cut plane by applying a random transform  $T_{GT}$  to  $T_{init}$  in a range of  $\pm 10$  mm on translation and  $\pm 5^\circ$  in rotation around each coordinate axis. We deliberately limited the ranges to 10 mm in translation and  $5^\circ$  in rotation for several reasons. these values are quite realistic compared to the possible positions of the endoscope tip, and also this avoids sampling planes at the edges of the volume where there is no informative image data because it is outside the CT imaging cone, Then we simulate the US images from this oblique cut-plane to have the fixed image  $I_f$  using the simulation method of [3] described in section (2.4.4.2).

Finally, the training sample is represented by  $(V_m; I_f; T_{GT})$ . And the intensities of these two volumes/images are scaled between  $[0, 1]$ .

At the end we had 126 samples for training and 54 samples for testing. During the training the predicted transformation parameters will be relative to the initial transform. And the final estimated transform will be calculated through matrix manipulation.

$$T_{Fin} = T_{init} \oplus T_{Est} \quad (3.5)$$

Since the networks use voxels (and not mm), the translation parameters are expressed in voxels. At the end the estimated translation components can be simply scaled to their original size in mm. The rotational components remain unchanged.

The method is implemented using Pytorch. Optimization is carried out for 10,000 iterations using the Adam algorithm with learning rate=0.001, and batch size of 16, which took approximately 18hours.

#### 3.5.4. Network training

During training, the registration network estimates the 6 dof of the transformation parameters, the loss function will be the L2 norm error between the estimated transform  $T_{Est}$  and the GT transform  $T_{GT}$ .

$$L = \|T_{GT} - T_{EST}\|^2 \quad (3.6)$$

we also used an unsupervised image similarity loss, after estimating the transformation parameters, an arbitrary image plane is randomly sampled from the 3D sub-volume by

applying the estimated transformation  $T_{EST}$  to the 3D sub-volume, using the rigid grid generator and the resampler. The rigid grid generator takes the estimated parameters  $T_{Est}$  as input and generates a transformed resampling grid which has the same size height and width as the moving image  $V_m$  ( $H \times W$ ). By applying bilinear interpolation at each point location defined by the sampling grid, the resampler can get the intensity at a particular pixel in the wrapped image, this loss can significantly reduce the registration error.

The final training loss function can be formulated as sum of three losses: 1) the L2 norm of the error between the  $T_{GT}$  and predicted transformation parameters  $T_{Est}$ ; 2) an image similarity loss is the Mutual Information (MI) between the fixed US image and the estimated transformed CT image plane; and 3) we added also a third term for our loss function which is MSE between the extracted CT image plane  $I_{GT}$ , and the transformed moving CT image which is improved the accuracy. The final loss function will be the weighted sum of three terms:

$$L = \alpha \|T_{GT} - T_{EST}\|^2 + \beta MI(I_f, I_{Est}) + \gamma MSE(I_{GT}, I_{Est}) \quad (3.7)$$

Depending on the values of the wights gives to each term of the losses  $\alpha$ ,  $\beta$ ,  $\gamma$ , the optimal prediction of  $T_{Est}$  was for  $\alpha = 0.4$ ,  $\beta = 0.2$ ,  $\gamma = 0.4$ .

### 3.5.5. Experiments and results

Very few papers have used a learning-based approach for slice-to-volume registration, and more specifically we couldn't find any work done previously for our specific application. We therefore decided to compare our results to the classical iterative method using the simpleITK library. We compared the results of our framework to the classical one in terms of registration accuracy and computation time.

for SimpleITK, we used MI as the similarity metric because of our multimodal case, and adaptive gradient descent as the optimizer.

Three comprehensive criteria were used for evaluation: the transformation estimation errors, the plane distance error and the registration computation time.

#### 3.5.5.1. Transformation estimation errors

We compared the parameters of the transformation obtained by the two methods (our network-based method and the classical iterative method) with that of the GT. A transformation is composed by a translation vector  $t$  ( $t_x, t_y, t_z$ ) and a rotation  $R$  described by its Euler angles ( $\vartheta_x, \vartheta_y, \vartheta_z$ ). We evaluate separately the translation errors and the rotation errors between the estimated pose of each of the 54 ( $18 \times 3$ )-validation image/sub-volume pairs and their associated GT (see section 3.4.4.2).

Figure 3.16 compares our results in a quantitative way with those obtained by the classical iterative method.

Using the iterative method presented in [77], we obtained mean errors of (0.794, 0.893, 0.922) degree for rotation, and translation parameters error (1.618, 1.8289, 1.875) mm, and standard deviation equal to (0.393, 0.581, 0.648) degree and (1.102, 1.234, 1.319) mm for the rotation and translation parameters respectively. Results are presented in

Table 3-1. The median error was (0.71,0.62,0.82) degree for the rotation, and (1.41, 1.72, 1.81) for the translation parameters as shown in Figure 3.16.

With our network-based method, we measure the error between the estimated transformation parameters  $T_{Est}$  and the ground truth  $T_{GT}$ . The median error was (0.63, 0.56, 0.78) degree for the rotation, and (1.29, 1.42, 1.71) for translation parameters as shown in Figure 3.16.

The mean error was (0.684, 0.706, 0.776) degree for the rotation, and (1.556, 1.695, 1.739) mm for the translation parameters, with a standard deviation of (0.423, 0.414, 0.416) degree, and (1.099, 1.202, 1.106) mm respectively as presented in Table 3-1.

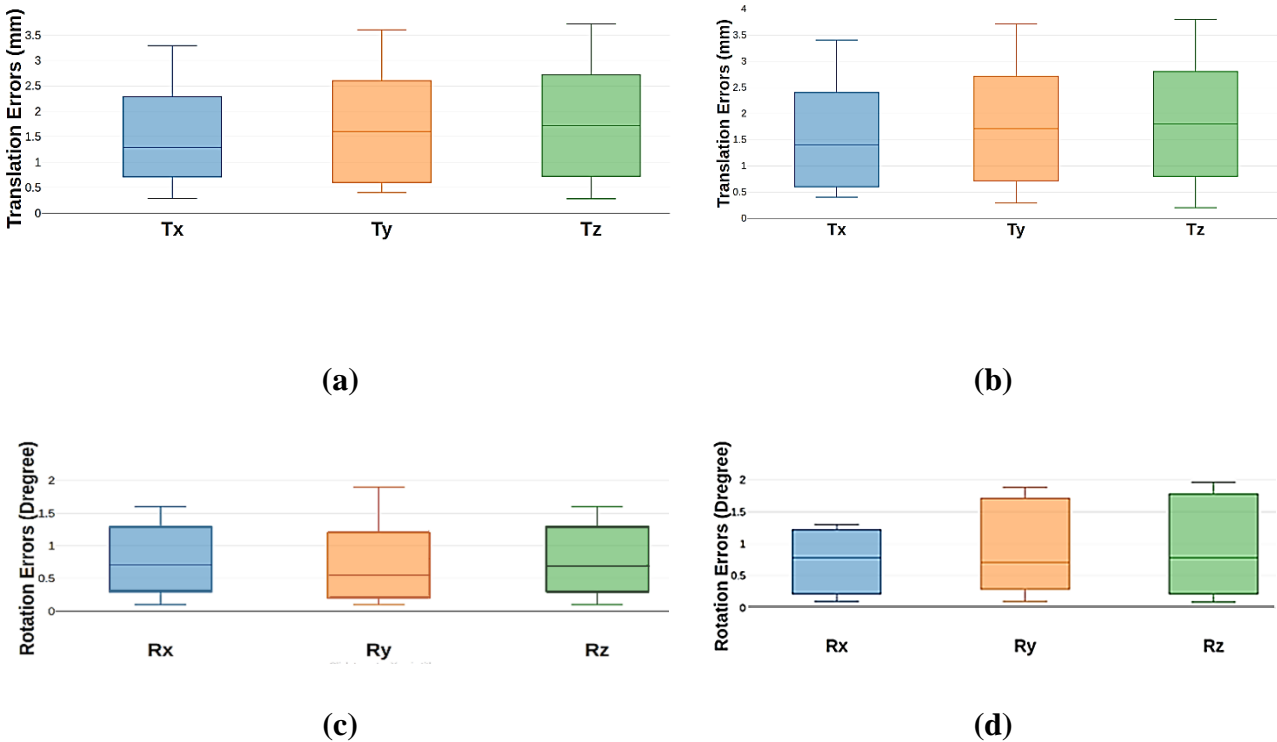


Figure 3.16 – Comparison of the error estimation for estimated rigid parameters (Rx, Ry, Rz), and (Tx, Ty, Tz) for our proposed method (Figures (a) and (c)) and the classical iterative approach presented by [77]. (Figures (b) and (d)).

### 3.5.5.2. Distance errors (DisErr)

The distance error denotes the average distance in millimeters between the oblique cut plane which is support of the ground truth  $I_{GT}$  (the ground truth image that we used to simulate the fixed US image), and oblique cut plane predicted by the parameter estimated by our method or the iterative one. The smaller the distance, the more accurate the estimation is. Table 3-1 shows a somewhat better performance (1.67 mm) has been achieved with our method compared to the iterative registration (1.89 mm).

### 3.5.5.3. Registration computation time

The average running time was around 0.070 second with our method compared to the classic iterative methods which takes around 28 seconds for one image/sub-volume pair. The results are presented in Table 3-1.

In fact, for classical methods, there are multiple factors that affect the computation time, (i.e., the number of pyramid resolution, the number of iterations, the number of grey level bins in each resolution level, the size of the optimization step ...). For example, using one resolution with 250 iterations takes around 9.75 seconds, compared to 4 multiresolution levels that takes around 67 seconds with the same number of iterations.

In our case, were there is no large deformation, a single resolution is sufficient. The maximum number of iterations was 250 iterations, we used adaptive step size with maximal size of 0.1, and the number of histogram bins is 32 for the MI.

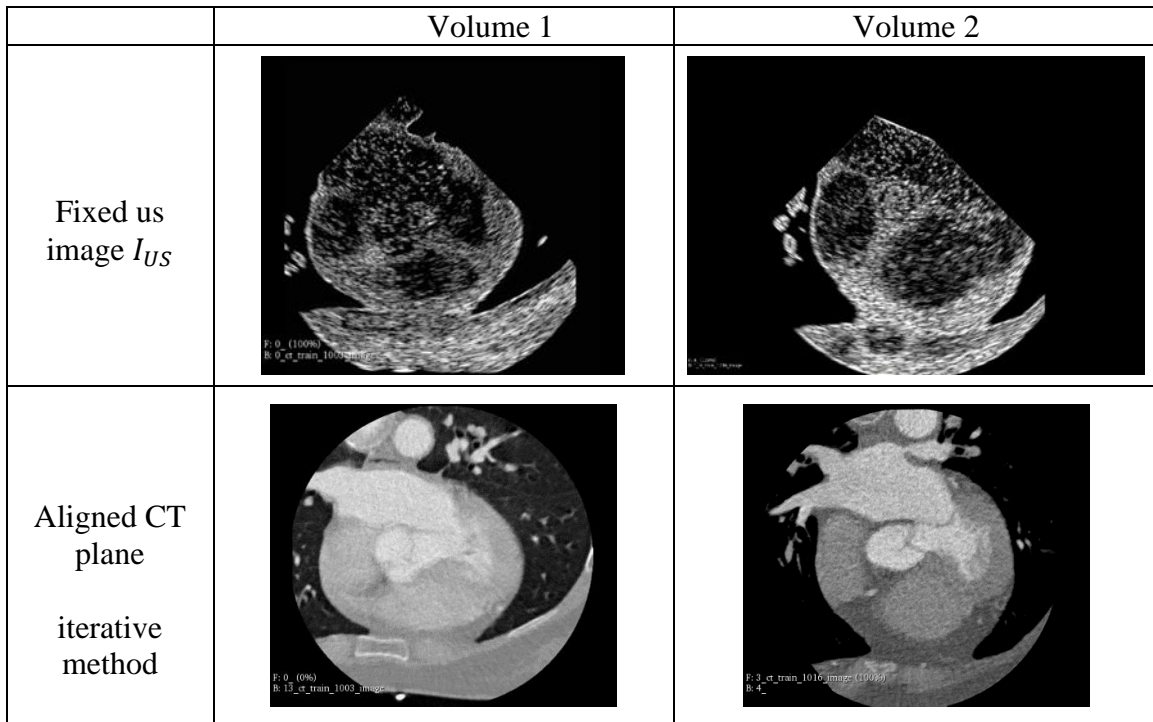
Table 3-1 – Performance comparison of our proposed framework and the classical iterative method.

Method	DisErr (mm)	Transformation Estimation Error							Time (Sec)
			$t_x$	$t_y$	$t_z$	$\vartheta_x$	$\vartheta_y$	$\vartheta_z$	
Iterative method	1.89	Mean	1.618	1.8289	1.875	0.794	0.893	0.922	9.65
		SD	1.102	1.234	1.319	0.393	0.581	0.648	
Our framework	1.67	Mean	1.556	1.695	1.739	0.684	0.706	0.776	<b>0.07</b>
		SD	1.099	1.202	1.106	0.423	0.414	0.416	

### 3.5.5.4. Visual validation

Figure 3.17 shows the overlap between the source image and the corresponding target plane, after registration using the classical iterative method, and our proposed learning-based method.

As we can qualitatively observe, the overlap increases after registration with better alignment when using the proposed method.









<p>Aligned CT plane</p> <p>Our framework</p>	 <p>F: 0_ (0%) B: 1_ct_train_1003_image</p>	 <p>F: 4_ (0%) B: 1_ct_train_1016_image</p>
<p>US superimposed on the aligned CT plane</p> <p>iterative method</p>	 <p>F: 0_ (40%) B: 13_ct_train_1003_image</p>	 <p>F: 3_ct_train_1016_image (0.99) B: 4_</p>
<p>US superimposed on the aligned CT plane</p> <p>Our proposed method</p>	 <p>F: 0_ (50%) B: 1_ct_train_1003_image</p>	 <p>F: 4_ (50%) B: 1_ct_train_1016_image</p>
	<p>Volume1</p>	<p>Volume2</p>

Figure 3.17 – examples of the registration results.

### 3.5.6. Discussion

In this section we present a supervised learning-based registration approach to perform slice-to- volume learning-based registration for cardiac arrhythmia guidance therapy, by using CNN to assess non iteratively the rigid transformation parameters between an US slice and the CT volume.

Every registration case took around 0.07 seconds (almost 140 times less than the classical method). This acceleration was not obtained at the expense of the registration accuracy because with our method, this accuracy is of the same order of magnitude or even slightly better.

The experimental results demonstrate the equally good registration performance of our work and especially a much higher computation speed compared to the conventional iterative registration method.

### **3.6. Conclusion**

In this chapter, we first discussed the evolution of medical image registration methods to deep learning-based approaches. We presented then the existing methods and some potential directions for future research like GAN. We gave a comprehensive summary of published papers focused on supervised deep learning-based medical image registration algorithms for both mono and multimodal imaging. We also highlighted one of the most challenging problems in medical image registration: the Slice-To-Volume registration for both projective and extracted slices using learning registration approaches.

Then, we presented our framework for the registration of US CT image pairs. We propose the following process: we first ran the input moving and fixed image pairs through a siamese architecture composed of convolutional layers which was able to extract features of the moving and fixed images analogous to dense local descriptors, then we matched the two feature maps, and finally we ran this corresponding feature maps into a registration network, which directly gave as output the registration parameters set of the rigid registration. The accuracy of the registration has been quantified based on the Target Registration Error (TRE) for specific anatomical landmarks. Results of the registration process showed a median TRE of 2.2 mm for all the fiducial points, and the registration computation time was around 3 ms comparing to the classic iterative methods which took around 6 seconds for one image pair.

And finally, we extended our work to tackle the most challenging problem of 2D US to 3D volume image registration in order to refine the estimation of the transesophageal probe pose in the 3D preoperative volume, which is essential for real time image guidance therapy. The proposed framework consists of two branches of CNN to extract the feature maps of the fixed US image and the moving CT volume, then we combined the two feature maps into a concatenation layer, and finally we passed these features into ResNet 50 to assess the six rigid transformation parameters.

The evaluation showed the proposed method is able to achieve state of the art results in terms of accuracy, while decreasing the computation time compared to the classical iterative methods method. In this multimodal case we were able to reduce the computation time from around 9.65 sec to 0.07 sec. This new method will allow us to find the pose of the TEE US image in the CT volume in a more precise way and especially with a computation time compatible with the clinical routine.

For this work, because of the imaging probe is still under development, we couldn't have real US images, also we didn't have enough contact with doctors to provide us numeric phantom, or real images because of COVID19 conditions that's covered the most duration of the work.



## Chapitre 4

# Learning-Based Registration: unsupervised Transformation Estimation

### 1.1. Introduction

Despite the success of the supervised methods, the difficult nature of the acquisition of reliable ground truth remains a significant obstacle in real data [117]. This has motivated a number of different groups to explore unsupervised approaches for image registration.

Currently, unsupervised methods are the hot topic in medical image registration, as they can predict the deformation fields and warped moving images in a single pass, and do not require ground-truth transformations for training. Similar to supervised method.

The study presented in this chapter aims at quantifying the nonrigid US/CT registration. Thus, the deformations of the heart that result from the patient's breathing and the movement of the heart will be taken into account during the interoperable procedure.

The first part of this chapter, we present a review of the literature on the use of unsupervised deep learning approach in the medical domain and their applications, then we present our proposed framework, and finally our results obtained on real and simulated datasets.

### 4.1. Background

As demonstrated in the previous chapter, several CNN architectures have been proposed in recent years, such as AlexNet [84], VGG [85], ResNet [86], and DenseNet [87]. Among these, in the area of medical image segmentation and registration, the most widely used architecture is the U-Net [88]— an encoder–decoder style network with skip connections between the encoding and decoding paths (as depicted in Figure 4.1). The encoder contains several convolutional layers and pooling layers, which downsample the input image to a low resolution. While the decoder is made up of deconvolution layers with a matching number of layers to the encoder. Through the decoder, the feature maps are reconstructed to the original size of the input images.

The U-Net utilizes multiple down- and up-sampling layers to learn features at different resolutions, with limited expense of computational resources. It has been widely applied in various medical imaging applications (*e.g.*, segmentation), and due to its flexibility, most state-of-the-art Deep Learning-based medical Image Registration methods use it, as well as in some component of the overall framework, where the final fully-connected layer can be dropped out, so that a direct end-to-end registration field can be achieved.



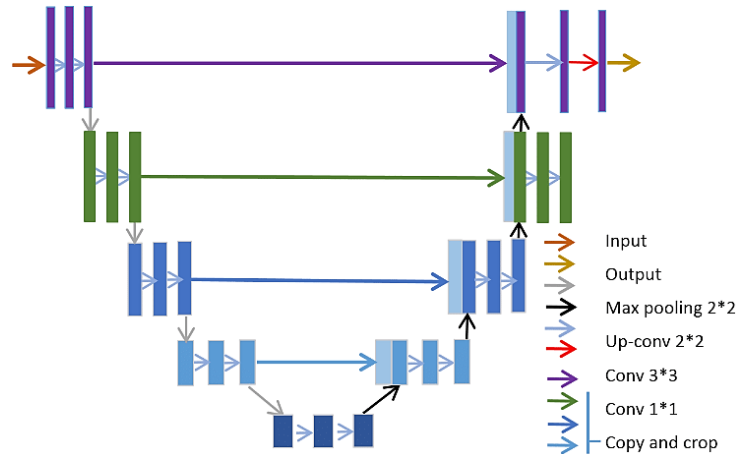


Figure 4.1 – An example of the U-Net architecture.

Currently, unsupervised non rigid image registration methods can predict the deformation fields and warped moving images in a single pass, and do not require ground-truth transformations for training.

Since no ground-truth data is available or used, the first problem to tackle in training unsupervised registration networks, is to formulate a loss function that can be optimized to train the network. Using a spatial transformer, deep learning networks can generate some deformation fields to warp the moving image. The dissimilarity between the warped moving image(s) and fixed image(s) can then be used to calculate the loss function for back-propagation. This measure of dissimilarity (or similarity) is typically estimated using an image similarity metric such as Mean Square Error (MSE) and Mutual Information (MI).

Convolutional neural networks (CNNs) form the basis for most deep learning-based image registration networks. In 2017, De Vos et al [118] were the first to propose an unsupervised end-to-end network, based on CNN, to register 2D cardiac cine MRI images. They demonstrated that the registration accuracy of their approach was comparable to SimpleElastix. Similarly, Jun et al [119] proposed a ‘CNN’ network for the registration of 2D abdomen MRI, which was the first CNN-based registration method for abdominal images.

## 4.2. CNN in deformable medical image registration: Related work

As discussed in section 3.3 the fundamental blocks for image registration remain the same in either deep learning or iterative based approaches, and the role of deep learning part is to enhance or replace one of the components of the registration framework (metric, transformation model) or to facilitate other operations such as features extraction. In our case we will use CNN to directly estimate the transformation parameters from the input image pair.

In the rest of this bibliographic section, we will focus on the use of learning-based approaches in deformable (non-rigid) medical image registration.

### 4.2.1. Deep Similarity based Registration

In this section, we review some methods that use deep learning to learn a similarity metric. This similarity metric is then inserted into a classical intensity-based registration framework with a defined interpolation strategy, transformation model, and optimization algorithm. In this scenario deep neural network acts as an approximator of the similarity between the input images. This complete and no faulty similarity metric can then be inserted into the registration process.

**Related works:** Simonovsky et al proposed a 3D similarity network using a few aligned image pairs [120]. The network was trained to classify whether an image pair is aligned or not. They observed that the hinge loss performed better than the cross-entropy loss. The learned deep similarity metric was then used to replace MI in the traditional deformable image registration (DIR) for brain T1-T2 registration. It is important to ensure the regularity of the first order derivative in order to adapt the deep similarity metrics to the traditional deep image registration frameworks. The gradient of the deep similarity metric with respect to transformation was calculated using chain rule. They found that high overlap of neighboring patches led to smoother and more stable derivatives. They trained the network using the IXI brain dataset and tested it using a completely independent dataset called ALBERTs to show the good generality of the learned metric. They showed that the learned deep similarity metric outperforms MI by a significant margin. Compared with CT-MR and T1-T2 image registration, MR-US image registration is more challenging due to the fundamental imaging differences in image acquisition principles between MR and US.

### 4.2.2. Unsupervised Transformation Estimation

The underlying philosophy behind these approaches is that the deep neural network acts as a regressor to directly estimate the transformation parameters in a single run.

Instead of using a huge ground truth set, we can use data augmentation techniques on a small numbers of input samples as a traditional similarity measure (or combination thereof) can then be used as a loss function to guide the learning process, to maximize speed of execution. This last point is critical in a real-time application such as transesophageal HIFU image guidance therapy.

**Related works:** Uzunova et al. [117] generated ground truth data using statistical appearance models (SAMs). They used CNN (an adaptation of FlowNet to estimate the deformation field for the registration of 2D brain MRs and 2D cardiac MRs. They demonstrated that training FlowNet with ground truth data generated by SAMs yielded superior performance to CNNs trained with randomly generated ground.

One of the important components of most of the unsupervised deep learning image registration approaches is the spatial transformer network (STN), proposed in 2015 [121], STN learns to spatially transform feature maps in a way that is beneficial to the task of interest. Although not explicitly designed for image registration, but rather to imbue networks with the means to learn features in a way that is invariant to rigid and deformable transformations, they have become the basis for most unsupervised registration methods. The STN consists of three components: a localization network, a

grid generator and a sampler. The localization network is a CNN, which takes feature maps as input and outputs the parameters of a suitable/user-specified spatial transformation. The transformation parameters are then used to generate a resampling grid by the grid generator. Finally, a differentiable image sampling is performed by a linear sampler using the grid generated of the previous step.

Balakrishnan et al proposed an unsupervised deep image registration method for MR brain atlas-based registration [122]. Their approach was based on a ‘U-Net+STN’ framework with different traditional similarity metrics (MSE and CC) for 3D brain MRI image registration. They used a U-Net like architecture and named it ‘VoxelMorph’. During training, the network penalized the differences in image appearances with the help of the spatial transformer network. A smoothing constraint was used to penalize local spatial variations in the predicted transformation. They achieved comparable performance to the ANTs [123] registration method in terms of Dice Score Coefficients (DSC) of multiple anatomical structures. Subsequently, they extended their method to exploit auxiliary segmentations available in the training data. A DSC loss function was added to the original loss functions in the training phase. Segmentation labels were not required during testing. They investigated unsupervised brain registration, with and without DSC loss on the segmentation label. Their results showed that the segmentation loss contributed to improve the DSC scores. The performance is comparable to ANTs and NiftyReg [124], while being x150 faster than ANTs and x40 faster than NiftyReg.

Similar to Balakrishnan et al [122], Qin et al also used segmentation as complementary information for cardiac MR image registration [125]. They found that the features learned by CNN registration could also be used in segmentation. The predicted Deformation vector fields (DVF) were used to warp the masks of the moving image to generate the masks of the fixed image. They trained a joint segmentation and registration model for cardiac cine image registration and proved that the joint model could generate better results than the two separate models alone in both segmentation and registration tasks.

Later, Zhang proposed a network with trans-convolutional layers for an end-to-end prediction of the DVF in MR brain DIR [126]. They focused on the diffeomorphic mapping of the transformation. To encourage smoothness and avoid folding of the predicted transformation, they proposed an inverse-consistent regularization term to penalize the difference between two transformations from the respective inverse mappings. The loss function consists of an image similarity loss, a transformation smoothness loss, an inverse consistent loss and an anti-folding loss. Their method outperformed the Demons and symmetric normalization metrics, in terms of DSC score, sensitivity, positive predictive value, average surface distance and Hausdorff distance [127].

A similar idea was proposed by Kim et al who used cycle consistent loss to enforce the regularization of the DVF [128]. They also used identity loss where the output DVF should be zero if the moving and fixed image are the same.

Rohe et al. [129] also used a network inspired by U-net [88] to estimate the deformation field used to register 3D cardiac MR volumes. Mesh segmentations are used to compute the reference transformation for a given image pair and the SSD between the prediction

and the ground truth is used as loss function. This method outperformed LCC Demons based registration [130].

In another work, Vos et al. [118] used NCC to train an FCN to perform the deformable registration of 4D cardiac cine MR volumes. A DVF is used in this method to deform the moving volume. Their method outperforms registration that is performed using the Elastix toolbox [131]. Indeed, this work should be considered as the most comprehensive deep learning framework based on the CNN to directly estimate the deformable transformation parameters in a single shot. In addition to estimate the transformation parameters, their multi-stage multi-resolution approach was able to learn a predefined similarity measure so that the need to use a synthesized and labeled dataset is avoided, which is a great advancement in the application of CNNs to the field of medical image analysis where we are faced with small-sized annotated datasets.

Sun et al. [132] proposed an unsupervised method for 3D MR/US brain registration that uses a 3D CNN consisting of a feature extractor and a deformation field generator. This network is trained using a similarity metric that incorporates both pixel intensity and gradient information. In addition, both image intensity and gradient information are used as inputs to the CNN.

Finally, for 3D-CT image registration, Hering et al [133] combined three 2D networks to construct a 2.5D registration approach, for cardiac MRI-CT registration. They demonstrated that their approach achieved a higher Dice score than previous state-of-the-art unsupervised registration methods.

### 4.3. Deformable US/CT registration with a convolutional neural network

In this section, we present our proposed approach, a Convolutional Neural Network (CNN) framework for deformable transesophageal US/CT image registration which will be used for cardiac arrhythmias and guidance therapy purposes. This framework consists of three main parts (Figure 4.2): a) a CNN, b) a spatial transformer and c) a Resampler. CNN receives concatenated pairs of moving and fixed images as input and estimates the parameters of the spatial transformer as output. The spatial transformer generates the displacement vector field allowing the resampler to wrap the moving image in the fixed image.

In our approach, we train the model to maximize standard image matching objective functions that are based on image intensities. The network can be applied to perform non-rigid registration of a pair of CT/US images directly in a single pass, thus avoiding the time-consuming computation load of the classical iterative method.

In this work, we focus on the registration of a 2D CT slice to a transesophageal 2D US image with an unsupervised learning approach. The network can be applied to perform the registration on unknown image pairs in single pass, thus in a non-iterative manner to avoid time consuming issues. This approach should also allow elastic registration which is generally more suited to handling cardiac images.

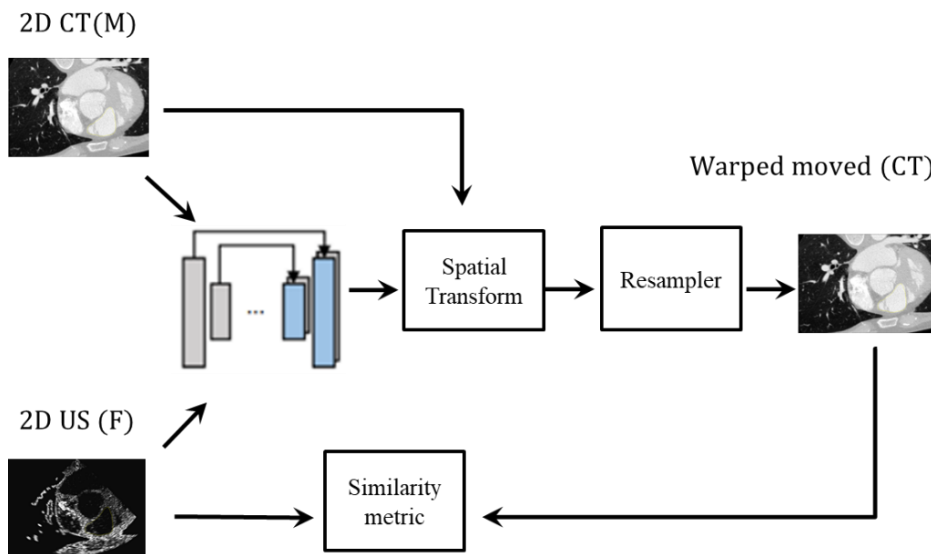


Figure 4.2 – The general framework of the proposal approach.

#### 4.3.1. CNN model

The network architecture is similar to that of the U-Net [88], which consists of encoding and decoding sections with skip connections. The convolutional layers of the encoder capture hierarchical features of the input image pair that are then used to estimate the DVF in the decoding stage. The CNN receives concatenated pairs of moving and fixed images as input and applies two alternating layers of  $2 \times 2$  convolutions in both the encoder and decoder stages using a kernel size of 3, each of which is followed by a rectified linear unit (ReLU) and a  $2 \times 2$  max pooling operator with a stride of 2 down-sampling layers in the encoder path to reduce the number of the CNN parameters. Each decoding step consists of an up-sampling, convolutions (“up-convolution”) that halve

the number of feature channels, The skip connections propagate the features learned during the encoding stages directly to the layers generating the registration. This enables the alignment of the image pairs.

The network is trained by optimizing an image similarity metric (i.e. by dissimilarity backpropagation ) between pairs of moving and fixed images of a training set using the ADAM optimizer [134]. After training, the network can be applied for unseen images registration.

We implemented the network using Tensorflow and trained it on a NVIDIA TitanX GPU with 1000 iterations which took approximately 6 hours. After training, the network can be applied for registration of unseen images.

### 4.3.2. Spatial transform

The spatial transformer generates the DFV that enables the resampler to wrap the moving image in the fixed image. The spatial transform is based on the spatial transformer network [122]. It computes for each pixel  $p$ , the new location in the warped moving image by adding the displacement vector  $(dx, dy)$  to that pixel.

Since mapping from one space to the other will often require an estimation of the intensity of the image at non-grid positions, an interpolator is required.

### 4.3.3. Loss function

We use mutual information (MI) as our loss function. In a previous case-based test we found that MI was one of the best-fitting similarity measure for our US/CT data [4]. MI compares the information of the US images and the corresponding information extracted from the CT slices.

$$L(A, B) = \sum_{a, b} p(a, b) \log \frac{p(a, b)}{p(a)p(b)} ; a \in A, b \in B \quad (8)$$

Where  $A$  is the reference image (US),  $B$  represents the warped moving image (transformed CT slice),  $p(a)$ ,  $p(b)$  are the marginal distributions of  $A$  and  $B$ , and  $p(a, b)$  their joint distribution, we also added a regularization term to energy to make the deformation smooth or more realistic.

## 4.4. Experiments and results

In order to produce enough data for the training and testing of our framework and also to enhance the robustness of the method, we arbitrary produced 250 poses in a range of  $\pm 5$  mm on translation and  $\pm 5^\circ$  in rotation around an initial pose located in the centerline of the esophagus in the middle of the volume, and we extract all the 2D CT slices using the framework described in [135]. We then simulated the corresponding US images with the method described in [3] (see section 2.1.4.2).

Every CT and US image pair is resampled to the same resolution according to the spacing information and overlap the moving and fixed image with each other by applying an initial transform.

transform to from this dataset, the network was trained by randomly selecting 175 pairs of fixed (simulated US images) and moving (cardiac CT plane) image slices. The 75 remaining pairs were used for validation.

The pairs of fixed and moving images were anatomically corresponding slices, but we added some artificial deformations when we simulated the 2D US images. To deform the images, we used the Elastodeform python library. The idea was to generate a coarse displacement grid with a random displacement for each knot of this grid. The image is then deformed using these displacement vectors and a spline interpolation.

#### 4.4.1. Dataset

This study has been conducted on a patient with cardiac fibrillation CT dataset, obtained from Louis Pradel University Hospital in Lyon, France, and simulated US images. The dimensions of the reconstructed CT volume are  $512 \times 512 \times 323$  voxels with an image spacing of  $0.546875 \times 0.546875 \times 0.55031$  mm<sup>3</sup>.

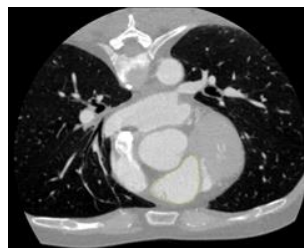
#### 4.4.2. Experimental protocol

We compared the registration results obtained by the proposed methods to those obtained by an iterative B-Spline free form deformation field non-rigid registration method implemented in the SimpleElastix Library [136]. We used the same similarity measure MI in both methods.

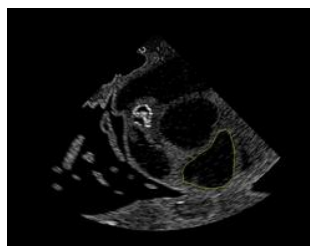
#### 4.4.3. Qualitative visual evaluation

Figure 4.3 shows an example of the registration of an image pair: a) The slice extracted from the CT volume; b) The simulated US image; c) The overlap between the moving CT image and the fixed US image using the classical method from SimpleElastix and d) our CNN one.

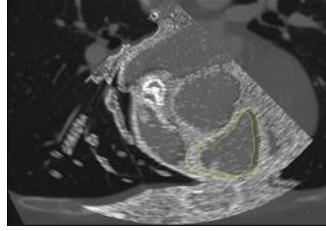
Visually, the results obtained by the proposed method seem to provide a better alignment than the classical free-form deformation field method. This can be seen for example at the bottom of the image on the chest.



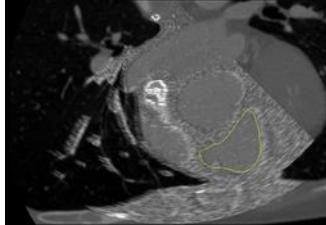
(a)



(b)



(c)



(d)

Figure 4.3 – (a) 2D CT slices, (b) simulated 2D US slices with superimposed boundaries of the left atrium (yellow). The deformed moving images obtained by (c) the free-form deformation field method of SimpleElastix and (d) the proposed approach.

#### 4.4.4. Quantitative evaluation

The comparison of the accuracy of the registration was performed according to two complementary metrics: the Dice similarity coefficient and the Hausdorff distance.

##### 4.4.4.1. Dice similarity coefficient (Dsc)

Obtaining some dense ground truth registration for this kind of data is not easy because many registration fields can produce similar-looking warped images. For this reason, we evaluated our method using volume overlap of anatomical segmentations.

As a first step, we selected anatomical structures that have a volume of at least 100 voxels for all tested subjects. The ideal candidate structure in our case was the left atrium. We then manually segmented all test image pairs (400 images) using ITKsnap.

The Dice similarity score is a widely used non-parametric measure to quantify the amount of overlapping regions between the input fixed image and the warped moving images, It computes the number of pixels that overlap between two surfaces and normalizes it by the half of the sum of the number of non-zero pixels in the two surfaces [137]:

$$\alpha = \frac{2|A \cap B|}{|A| + |B|} \quad (9)$$

Where  $A$  is the ground truth standard surface which, in our case, refers to the segmentation of the left atrium in the US fixed image (see Figure 4.3), and  $B$  the segmentation extracted from the warped registered CT image.

$\alpha = 1$  indicates that the anatomy matches perfectly after registration, and  $\alpha = 0$  indicates that there is no overlap. If a registration correctly estimates the precise anatomical correspondence between the two images, we expect that the regions in the fixed and in the moving image that correspond to the same anatomical structure will overlap well.



#### 4.4.4.2. Hausdorff distance

The Hausdorff distance [127] is a metric based on spatial distance. It is defined as the maximum of the nearest distance between two objects where the nearest distance is computed for each point or vertex of the contour of the two objects. A smaller Hausdorff distance indicates a closer topology between the two objects. These two metrics were measured on the boundaries of the surface of the segmented left atrium (see Figure 4.3).

**Error! Reference source not found.** compares the measurement indices obtained on the 400 test images, namely: the average Dice similarity coefficient, the average Hausdorff distance and the average computation time. These indices were obtained by using the iterative method of SimpleElastix and by our CNN (U-Net) based method. As a reminder, the Dice measurement and the Hausdorff distance were conducted on the segmented left atrium. In each case we report the mean and standard deviation of the measured scores on the validation data set.

As expected, we can see in Table 4-1 that the computation time is greatly improved using CNN (under a second) than using the classical iterative method (around one minute). More surprising, we can also notice in Table 4-1 that both spatial comparison metrics are improved using CNN compared to the classical approach.

Table 4-1– Average Dice similarity coefficient, Hausdorff distance and computation time results for SimpleElastix and CNN (U-Net) across the segmented left atrium in the US fixed image, and the segmentation mapped from the warped registered CT (see Figure 4.3).

Method			
	<i>Dice sim. Coef.</i>	<i>Hausdorff distance (mm)</i>	<i>Comp. time (sec)</i>
SimpleElastix	0.7 (0.01)	1.7 (0.02)	65 (0.1)
CNN (U-Net)	0.8 (0.02)	1.2 (0.05)	0.7 (0.02)

These results were obtained from simulated US images. We are well aware that there are differences between simulated US images and real US images. However, we have found in a previous study that a method tuned on simulated data allowed to have also good results on real data [7]. So, we are hopeful that our method works on real data.

#### 4.5. Conclusion

In this chapter, first we presented a review of unsupervised deformable image registration in the medical domain and their applications.

Then, we present our proposed unsupervised learning-based-approach for transesophageal US/CT cardiac image registration. The obtained results indicate a strong improvement in terms of computation time without any loss (even with some improvements) in terms of registration accuracy.

In our future work, we will integrate the unsupervised learning approach as a second step after the 2D US/3D CT image rigid registration method proposed in Chapter 3 for

probe navigation inside the esophagus during the non-invasive cardiac arrhythmia ablation performed by real time HIFU.



# General conclusion and perspective

Minimally-invasive transesophageal HIFU ablation technique for cardiac arrhythmia therapy is a very promising treatment that could be the future direction for cardiac arrhythmia medicine. This therapy will improve the ablation procedure because it can go deeper in the tissues without damaging the intervening tissues, unlike the radiofrequency catheter ablation. This technique will also reduce the cost for treating cardiac arrhythmia.

One of the most challenging issues for this technology is to find image processing solutions that can accurately define the position of the TEE imaging probe inside the patient's body during the real-time treatment. These solutions should be both accurate and fast for real time guidance therapy.

The main objective of the presented work is to propose multimodal image registration solutions for image guidance therapy, to navigate the transesophageal HIFU probe with an embedded TEE imaging device inside the esophagus toward the arrhythmic area to treat it. For this we need to combine the information from the preoperative high resolution (CT/MRI) volume, and the intraoperative low resolution 2D US images to give physicians the best guidance information to define the lesion position. For this purpose, we have proposed and developed three solutions to register 2D ultrasound/3D CT volume based on classical and deep learning methods. Three main contributions can be highlighted in this work:

1. The first contribution we proposed was based on a classical iterative intensity-based registration of the two perpendicular 2D US to preoperative 3D CT. As a proof of concept we developed the following evaluation framework on a digital phantom: 1) as the probe is under development we defined a ground truth (GT) initial pose inside a CT volume and simulated two perpendicular US images from the CT data; 2) we ran the registration framework from 55 randomly defined initial pose around the initial GT pose; and 3) we estimated the accuracy of the registration by (a) the errors on the estimated transformation parameters (translation error and quaternion distance for rotation) and (b) Target Registration Error (TRE) on 8 features. The accuracy of the registration using two 2D US planes has been compared to the previous work using only a single US plane. An improvement was observed when using two 2D US planes compared to the previous single US plane. The median translation errors were reduced from 1,5 to 0,7 mm, the median rotation error from 3,2° to 2,1° and the median TRE from 2,5 mm to 1,76 mm.
2. In the second contribution, we presented a convolutional neural networks (CNNs) framework for transesophageal ultrasound/computed tomography image registration. The use of CNNs could potentially solve the problem of high computation time of the classical iterative methods, which are not suitable for a real-time application. We proposed the following process: we first ran the input

moving (CT) and fixed image pairs through a siamese architecture composed of convolutional layers, to extract features from the moving and fixed maps analogous to dense local descriptors; then we matched the feature maps; and finally ran this correspondence feature map through a registration network, which directly outputs the registration parameters set of the rigid registration. Accuracy of the registration is quantified based on the Target Registration Error (TRE) for soft specific anatomical landmarks. The results of the registration process showed a median TRE of 2.2 mm for all the fiducial points, and the registration computation time was around 3 ms compared to the classic iterative methods which took around 6 seconds for one image pair. This contribution confirmed that the network could detect common features between the CT and US image pairs.

We then extended this first 2D/2D registration study to propose a learning-based 2D/3D registration to refine the estimation of the transesophageal probe pose in the 3D preoperative volume. The proposed framework consisted of two networks to extract the feature maps of each pair of fixed US image and CT sub-moving volume, followed by a concatenation layer, and finally the registration network ResNet 50 was used to estimate the six rigid transformation parameters.

As we have shown, compared to a classical iterative method, the quality of the results was preserved (and improved in some cases) while the computational time was highly reduced. Every registration case took around 0.07 seconds (almost 140 times less than the classical iterative method).

3. Finally, in our third contribution, we presented a Convolutional Neural Network (CNN) framework for deformable transesophageal 2D US/2D CT image registration, for the cardiac arrhythmias therapy guidance. The framework consisted of a CNN, a spatial transformer, and a resampler. CNN received concatenated pairs of moving and fixed images as input and estimated the parameters for the spatial transformer as output. The spatial transformer generated then a displacement vector field that allowed the resampler to wrap the moving image in the fixed image. In our proposed method, we trained the model to maximize some standard image matching objective functions that are based on the image intensities. The network can be applied to perform non-rigid registration of a pair of CT/US images directly in a single pass, avoiding the time-consuming computation of the classical iterative method. The results compared the two approaches in terms of computation time and accuracy. The computation time was highly improved using CNN (under a second) compared to the classical iterative method (around one minute). We could also notice that both spatial comparison metrics are improved when using CNN compared to the traditional approach.

### **Perspectives:**

Throughout the proposed work, we faced a variety of scientific challenges and in particular the difficulty of obtaining a real database or even physical phantom datasets for validation. Also, in the context linked to the COVID-19 pandemic we didn't have enough contact with the doctors to provide us even physical phantom data. So, we decided to validate our frameworks with simulated US images.

But many of these difficulties were overcome with dedicated methods. At the end of this work, several perspectives can be identified:

- 1) To combine the two algorithms of rigid and deformable image registration into a single algorithm with two steps: first to define the 3D pose of the probe in 3D volume, second to have an accurate description of the ablation area by applying the deformable transform between the 2D US fixed image, and the resulted transformed moving 3D CT slice (the first step result).
- 2) to apply some generative learning approach like GAN to generate more ground truth data.
- 3) To validate the results on physical phantom or/and real patients' data. We hope that a network tuned on simulated data will work on real data.



## References

- [1] G. Haar and C. Coussios, “High intensity focused ultrasound: Physical principles and devices,” *Int. J. Hyperth.*, vol. 6736, 2009, doi: 10.1080/02656730601186138.
- [2] D. E. Rennes, D. De, and D. E. Rennes, “Planning and guidance of ultrasound guided High Intensity Focused Ultrasound cardiac arrhythmia therapy,” 2015.
- [3] J. L. Dillenseger, S. Laguitton, and É. Delabrousse, “Fast simulation of ultrasound images from a CT volume,” *Comput. Biol. Med.*, vol. 39, no. 2, pp. 180–186, 2009, doi: 10.1016/j.combiomed.2008.12.009.
- [4] Z. L. Sandoval and J. L. Dillenseger, “Evaluation of computed tomography to ultrasound 2D image registration for atrial fibrillation treatment,” *Comput. Cardiol. (2010)*, vol. 40, pp. 245–248, 2013.
- [5] X. Zhuang and J. Shen, “Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI,” *Med. Image Anal.*, vol. 31, pp. 77–87, 2016, doi: 10.1016/j.media.2016.02.006.
- [6] X. Zhuang, “Challenges and methodologies of fully automatic whole heart segmentation: A review,” *J. Healthc. Eng.*, vol. 4, no. 3, pp. 371–407, 2013, doi: 10.1260/2040-2295.4.3.371.
- [7] Z. Sandoval, M. Castro, J. Alirezaie, F. Bessiere, C. Lafon, and J. L. Dillenseger, “Transesophageal 2D ultrasound to 3D computed tomography registration for the guidance of a cardiac arrhythmia therapy,” *Phys. Med. Biol.*, vol. 63, no. 15, 2018, doi: 10.1088/1361-6560/aad29a.
- [8] A. Timmis *et al.*, “European Society of Cardiology: Cardiovascular Disease Statistics 2019,” *Eur. Heart J.*, vol. 41, no. 1, pp. 12–85, 2019, doi: 10.1093/eurheartj/ehz859.
- [9] E. J. Benjamin *et al.*, “Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association,” *Circulation*, vol. 139, no. 10, pp. e56–e528, Mar. 2019, doi: 10.1161/CIR.0000000000000659.
- [10] V. Fuster *et al.*, “ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation--executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Com,” *J. Am. Coll. Cardiol.*, vol. 48, no. 4, pp. 854–906, Aug. 2006, doi: 10.1016/j.jacc.2006.07.009.
- [11] E. Constanciel *et al.*, “Design and evaluation of a transesophageal HIFU probe for ultrasound-guided cardiac ablation: Simulation of a HIFU mini-maze procedure and preliminary ex vivo trials,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 60, no. 9, pp. 1868–1883, 2013, doi: 10.1109/TUFFC.2013.2772.
- [12] X. Yin, L. M. Epstein, and K. Hynynen, “Noninvasive transesophageal cardiac thermal ablation using a 2-D focused, ultrasound phased array: A simulation study,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 53, no. 6, pp.



- 1138–1148, 2006, doi: 10.1109/TUFFC.2006.1642512.
- [13] R. Bailey, “The Function of the Heart Ventricles,” 2021. [thoughtco.com/ventricles-of-the-heart-373254](https://thoughtco.com/ventricles-of-the-heart-373254).
- [14] “The Heart.” <https://med.libretexts.org/@go/page/12536>.
- [15] N. Naqvi, K. P. McCarthy, and S. Y. Ho, “Anatomy of the atrial septum and interatrial communications,” *J. Thorac. Dis.*, vol. 10, no. Suppl 24, pp. S2837–S2847, 2018, doi: 10.21037/jtd.2018.02.18.
- [16] P. A. Davlouros, K. Niwa, G. Webb, and M. A. Gatzoulis, “The right ventricle in congenital heart disease,” *Heart*, vol. 92, no. SUPPL. 1, pp. 27–38, 2006, doi: 10.1136/hrt.2005.077438.
- [17] I. Osztheimer and G. Fülöp, *Cardiovascular system*, vol. 3. 2010.
- [18] David S.ParkGlenn I.Fishman, “Cell Biology of the Specialized Cardiac Conduction System,” in *Cardiac Electrophysiology: From Cell to Bedside*, 2014, pp. 287–296.
- [19] R. Gordan, J. K. Gwathmey, L. Xie, R. Gordan, J. K. Gwathmey, and L. Xie, “Autonomic and endocrine control of cardiovascular function,” vol. 7, no. 4, pp. 204–214, 2015, doi: 10.4330/wjc.v7.i4.204.
- [20] “Recommendations for the standardization and interpretation of the electrocardiogram: Part I: The electrocardiogram and its technology: A scientific statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Cli,” 2007.
- [21] O. J. Bestetti RB, “The surface electrocardiogram: a simple and reliable method for detecting overt and latent heart disease in rats,” *Braz J Med Biol Res*, vol. 23, 1990.
- [22] A. P. Voorhees, H. Han, and E. Program, “Biomechanics of Cardiac Function,” *HHS Public Access*, vol. 5, no. 4, pp. 1623–1644, 2016, doi: 10.1002/cphy.c140070.Biomechanics.
- [23] A. Proclemer *et al.*, “How are European patients at risk of malignant arrhythmias or sudden cardiac death identified and informed about their risk profile : results of the European Heart Rhythm Association survey,” pp. 994–998, 2015, doi: 10.1093/europace/euv203.
- [24] M. R. Podrid PJ, “Epidemiology and stratification of risk for sudden cardiac death,” *Clin Cardio*, vol. 28, 2005.
- [25] B. M. P. H. Rudic B, Tülümen E, Liebe V, Kuschyk J, Akin I, “Sudden cardiac death : Epidemiology, pathophysiology and risk stratification,” vol. 42(2), pp. 123–131, 2017.
- [26] D. Andreu *et al.*, “Long-term benefit of first-line peri-implantable cardioverter – defibrillator implant ventricular tachycardia-substrate ablation in secondary prevention patients,” pp. 976–982, 2017, doi: 10.1093/europace/euw096.
- [27] D. Hidano *et al.*, “Ventricular fibrillation waveform measures and the etiology of cardiac arrest,” *Resuscitation*, vol. 109, pp. 71–75, 2016, doi:

- 10.1016/j.resuscitation.2016.10.007.
- [28] D. D. Salcido *et al.*, “Effects of intra-resuscitation antiarrhythmic administration on rearrest occurrence and intra-resuscitation ECG characteristics in the ROC ALPStrial,” *Resuscitation*, vol. 129, no. May, pp. 6–12, 2018, doi: 10.1016/j.resuscitation.2018.05.028.
- [29] C. C. Aepc *et al.*, “2019 ESC Guidelines for the management of patients with supraventricular tachycardia The Task Force for the management of patients with supraventricular tachycardia of the European Society of Cardiology (ESC),” pp. 655–720, 2020, doi: 10.1093/eurheartj/ehz467.
- [30] L. Blier *et al.*, “Ventricular Tachycardia Ablation versus Escalation of Antiarrhythmic Drugs,” pp. 111–121, 2016, doi: 10.1056/NEJMoa1513614.
- [31] T. R. Lin T, Ouyang F, Kuck K-H, “ThermoCool® SmartTouch® Catheter – The Evidence So Far for Contact Force Technology and the Role of VisiTag™ Module,” *Modul. Arrhythm Electrophysiol Rev*, pp. 3–44, 2014.
- [32] J.-U. Kluiwstra, Y. Zhang, P. VanBaren, S. A. Strickberge, E. S. Ebbini, and C. A. Cain, “Ultrasound Phased Arrays for Noninvasive Myocardial Ablation : Initial Studies,” pp. 1605–1608, 1995.
- [33] S. A. Strickberger, T. Tokano, J. A. Kluiwstra, F. Morady, and C. Cain, “Extracardiac Ablation of the Canine Atrioventricular Junction by Use of High-Intensity Focused Ultrasound,” pp. 203–208, 1999.
- [34] M. Carias and K. Hynynen, “The Evaluation of Steerable Ultrasonic Catheters for Minimally Invasive MRI-Guided Cardiac Ablation,” vol. 598, pp. 591–598, 2014, doi: 10.1002/mrm.24945.
- [35] T. H. E. Mechanism *et al.*, “The mechanism of lesion formation by focused ultrasound ablation catheter for treatment of atrial fibrillation,” vol. 55, pp. 647–656, 2010, doi: 10.1134/S1063771009040216.THE.
- [36] C. J. K. and K. K. H. Schmidt B., Antz M., Ernst S., Ouyang F., Falk P., “Pulmonary vein isolation by high-intensity focused ultrasound: first-in-man study with a steerable balloon catheter,” *Hear. Rhythm*, vol. 4(5), pp. 575–84, 2007.
- [37] K. Neven, A. Metzner, B. Schmidt, F. Ouyang, and K.-H. Kuck, “Two-year clinical follow-up after pulmonary vein isolation using high-intensity focused ultrasound (HIFU) and an esophageal temperature-guided safety algorithm,” *Hear. Rhythm*, vol. 9(3), pp. 407–13, 2012.
- [38] J. Ninet, X. Roques, R. Seitelberger, C. Deville, and L. Pomar, “Surgical ablation of atrial fibrillation with off-pump, epicardial, high-intensity focused ultrasound: Results of a multicenter trial,” vol. 130, no. 3, pp. 1–8, 2004, doi: 10.1016/j.jtcvs.2005.05.014.
- [39] F. D. and S. N. B. Werner J., Park E. J., Lee H., “Feasibility of in vivo Transesophageal Cardiac Ablation Using a Phased Ultrasound Array,” *Ultrasound Med Biol*, vol. 36(5), pp. 752–60, 2010.
- [40] G. Herman, *Fundamentals of Computerized Tomography: image reconstruction from projections. Advances in Computer Vision and Pattern Recognition*. 2009.

- [41] G. Gao, G. Penney, Y. Ma, N. Gogin, and P. Cathier, “Registration of 3D transesophageal echocardiography to X-ray fluoroscopy using image-based probe tracking,” *Med. Image Anal.*, vol. 16, no. 1, pp. 38–49, 2012, doi: 10.1016/j.media.2011.05.003.
- [42] N. B. and C. N. J. Brum, S. Catheline, “Quantitative shear elasticity imaging from a complex elastic wavefield in soft solids with application to passive elastography,” *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 62, pp. 673–685, 2015.
- [43] M. Alexandre Costet, M. Ethan Bunting, M. Hasan Garan, and M. Elaine Wan, “Validation of electromechanical wave imaging in a canine model during pacing and sinus rhythm,” *Hear. Rhythm*, vol. 13, pp. 2221–2227, 2016.
- [44] J. Huang, S. & Miller, “Catheter Ablation of Cardiac Arrhythmias,” *3rd Ed. edn Elsevier*, 2014.
- [45] J. Sra and S. Ratnakumar, “Cardiac image registration of the left atrium and pulmonary veins,” *Hear. Rhythm*, vol. 5, no. 4, pp. 609–617, 2008, doi: 10.1016/j.hrthm.2007.11.020.
- [46] K. Mcleish, D. L. G. Hill, D. Atkinson, J. M. Blackall, and R. Razavi, “A Study of the Motion and Deformation of the Heart Due to Respiration,” vol. 21, no. 9, pp. 1142–1150, 2002.
- [47] T. . Huang, X., Moore, J., Guiraudon, G., Jones, D.L., Bainbridge, D., Ren, J., Peters, “Dynamic 2D ultrasound and 3D CT image registration of the beating heart,” *IEEE Trans. Med. Imaging*, vol. 28, pp. 1179–1189, 2009.
- [48] L. Ibanez, W. Schroeder, L. Ng, and J. Cates, “The ITK Software Guide,” *ITK Softw. Guid.*, vol. Second, no. May, p. 804, 2005, [Online]. Available: <http://www.itk.org/ItkSoftwareGuide.pdf>.
- [49] T. S. Yoo, *Insight into images: principles and practice for segmentation, registration, and image analysis*. 2004.
- [50] G. M. and P. S. Frederik Maes, Dirk Vandermeulen, “Multimodality Image Registration by Maximization of Mutual Information,” vol. 16, pp. 187–198, 1997.
- [51] and D. H. Studholme, C., D. Hill, “An overlap invariant entropy measure of 3D medical image alignment,” *Pattern Recognit.*, vol. 32.1, pp. 71–86, 1999.
- [52] H. B. Johann Hummel, Michael Figl, Michael Bax and and W. Birkfellner, “2D/3D registration of endoscopic ultrasound to CT volume data,” *Phys. Med. Biol.*, vol. 53, pp. 4303–4316, 2008.
- [53] S. Miao, Z. J. Wang, and R. Liao, “A CNN Regression Approach for Real-Time 2D/3D Registration,” *IEEE Trans. Med. Imaging*, 2016, doi: 10.1109/TMI.2016.2521800.
- [54] K. . San Jos’e Est’epar, R., Westin, C.F., Vosburgh, “Towards real time 2D to 3D registration for ultrasound-guided endoscopic and laparoscopic procedures,” *Int. Surgery, J. Comput. Assist. Radiol.*, vol. 4, no. 549–560, 2009.
- [55] G. Gill, S., Abolmaesumi, P., Vikal, S., Mousavi, P., Fichtinger, “Intraoperative

- Prostate Tracking with Slice-to-Volume Registration in MR,” 2008.
- [56] J. Wolfgang Birkfellner, Michael Figl, Joachim Kettenbach, T. N. and H. Hummel, Peter Homolka, Ruđiger Schernthaner, and Bergmann., “Rigid 2D/3D slice-to-volume registration and its application on fluoroscopic CT images. Medical Physics,” *Med. Phys.*, vol. 34, no. 246, 2007.
- [57] M. Zakkaroff, C., Radjenovic, A., Greenwood, J., “Recovery of Slice Rotations with the Stack Alignment Transform in Cardiac MR Series,” in *British Machine Vision Conference (BMVC)*, 2012, pp. 35.1-35.11.
- [58] Xu, H., Lasso, A., Fedorov, A., Tuncali, K., Tempany, C., Fichtinger, G., “Multislice-to-volume registration for MRI-guided transperineal prostate biopsy,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, pp. 563–572.
- [59] A. Elen *et al.*, “Automatic 3-D breath-hold related motion correction of dynamic multislice MRI,” *IEEE Trans. Med. Imaging*, vol. 29, no. 3, pp. 868–878, Mar. 2010, doi: 10.1109/TMI.2009.2039145.
- [60] S. Osechinskiy and F. Kruggel, “Deformable registration of histological sections to brain MR images using a hybrid boundary-based slice-to-volume approach,” *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2011, pp. 4876–4879, 2011, doi: 10.1109/IEMBS.2011.6091208.
- [61] J. Wolfgang Birkfellner, Michael Figl, Joachim Kettenbach, T. N. and H. Hummel, Peter Homolka, Ruđiger Schernthaner, and Bergmann, “Rigid 2D/3D slice-to-volume registration and its application on fluoroscopic CT images,” *Med. Phys.*, vol. 34, p. 246, 2007, doi: 10.1118/1.2401661.
- [62] P. Markelj, D. Tomaževič, B. Likar, and F. Pernuš, “A review of 3D/2D registration methods for image-guided interventions,” *Med. Image Anal.*, vol. 16, no. 3, pp. 642–661, 2012, doi: 10.1016/j.media.2010.03.005.
- [63] C.-C. Lin, S. Zhang, J. Frahm, T.-W. Lu, C.-Y. Hsu, and T.-F. Shih, “A slice-to-volume registration method based on real-time magnetic resonance imaging for measuring three-dimensional kinematics of the knee,” *Med. Phys.*, vol. 40, no. 10, p. 102302, Oct. 2013, doi: 10.1118/1.4820369.
- [64] D. . Fei, B., Duerk, J.L., Wilson, “Automatic 3D registration for interventional MRI guided treatment of prostate cancer,” *Comput. Aided Surg.*, vol. 7, pp. 177–183, 2002.
- [65] Xu, L., Liu, J., Zhan, W., Gu, L., “A novel algorithm for CT-ultrasound registration,” *IEEE Point-of-Care Healthc. Technol.*, p. 101{104, 2014.
- [66] Liao, R., Zhang, L., Sun, Y., Miao, S., Chafd’Hotel, “A Review of Recent Advances in Registration Techniques Applied to Minimally Invasive Therapy,” *Trans. IEEE Multimedia*, vol. 15, pp. 983–1000, 2013.
- [67] X. Huang *et al.*, “Dynamic 2D ultrasound and 3D CT image registration of the beating heart,” *IEEE Trans. Med. Imaging*, vol. 28, no. 8, pp. 1179–1189, 2009, doi: 10.1109/TMI.2008.2011557.
- [68] L. Frühwald, J. Kettenbach, M. Figl, and J. Hummel, “A comparative study on manual and automatic slice-to-volume registration of CT images,” vol. 19, no. 11, pp. 2647–2653, 2010, doi: 10.1007/s00330-009-1452-0.A.

- [69] B. Fei, J. L. Duerk, D. T. Boll, J. S. Lewin, and D. L. Wilson, "Slice-to-Volume Registration and its Potential Application to Interventional MRI-Guided Radio-Frequency Thermal Ablation of Prostate Cancer," vol. 22, no. 4, pp. 515–525, 2003.
- [70] N. Birkfellner, W. Figl, M. Kettenbach, J. Hummel, J. Homolka, P. Scherthaner, R. and H. T., Bergmann, "Rigid 2D/3D slice-to-volume registration and its application on fluoroscopic CT images.," *Med. Phys.*, vol. 34, p. 246, 2007.
- [71] R. Lasowski *et al.*, "Adaptive visualization for needle guidance in RF liver ablation: taking organ deformation into account," in *Proc.SPIE*, Mar. 2008, vol. 6918, doi: 10.1117/12.769847.
- [72] R. Micu, T. F. Jakobs, M. Urschler, and N. Navab, "A new registration/visualization paradigm for CT-fluoroscopy guided RF liver ablation," *Lect. Notes Comput. Sci.*, vol. 4190 LNCS, no. 1, pp. 882–890, 2006.
- [73] S. Xu, G. Fichtinger, R. H. Taylor, and K. R. Cleary, "Validation of 3D motion tracking of pulmonary lesions using CT fluoroscopy images for robotically assisted lung biopsy," *Med. Imaging 2005 Vis. Image-Guided Proced. Disp.*, vol. 5744, p. 60, 2005, doi: 10.1117/12.594910.
- [74] R. Smolíková-Wachowiak, M. P. Wachowiak, A. Fenster, and M. Drangova, "Registration of two-dimensional cardiac images to preprocedural three-dimensional images for interventional applications," *J. Magn. Reson. Imaging*, vol. 22, no. 2, pp. 219–228, 2005, doi: 10.1002/jmri.20364.
- [75] H. X, "Rapid registration of multi modal images using a reduced number of voxels," *Proc. SPIE Med. Imag*, pp. 1–10, 2006.
- [76] D. Skerl, B. Likar, and F. Pernus, "A protocol for evaluation of similarity measures for rigid registration.," *IEEE Trans. Med. Imaging*, vol. 25, no. 6, pp. 779–791, Jun. 2006, doi: 10.1109/tmi.2006.874963.
- [77] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, "The design of simpleITK," *Front. Neuroinform.*, vol. 7, no. DEC, 2013, doi: 10.3389/fninf.2013.00045.
- [78] A. Fedorov *et al.*, "3D Slicer as an image computing platform for the Quantitative Imaging Network," *Magn. Reson. Imaging*, vol. 30, no. 9, pp. 1323–1341, Nov. 2012, doi: 10.1016/j.mri.2012.05.001.
- [79] R. Dalvi and R. Abugharbieh, "Fast feature based multi slice to volume registration using phase congruency.," *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, vol. 2008, pp. 5390–5393, 2008, doi: 10.1109/IEMBS.2008.4650433.
- [80] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bull.*, vol. 1, no. 6, pp. 80–83, 1945, [Online]. Available: [https://www-jstor-org.passerelle.univ-rennes1.fr/stable/3001968?origin=crossref#metadata\\_info\\_tab\\_contents](https://www-jstor-org.passerelle.univ-rennes1.fr/stable/3001968?origin=crossref#metadata_info_tab_contents).
- [81] P. Lang, M. W. A. Chu, D. Bainbridge, G. M. Guiraudon, D. L. Jones, and T. M. Peters, "Surface-based CT-TEE registration of the aortic root," *IEEE Trans.*

- Biomed. Eng.*, vol. 60, no. 12, pp. 3382–3390, 2013, doi: 10.1109/TBME.2013.2249582.
- [82] K. Cleary and T. M. Peters, “Image-guided interventions: technology review and clinical applications.,” *Annu. Rev. Biomed. Eng.*, vol. 12, pp. 119–142, Aug. 2010, doi: 10.1146/annurev-bioeng-070909-105249.
- [83] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes, “Medical image registration.,” *Phys. Med. Biol.*, vol. 46, no. 3, pp. R1-45, Mar. 2001, doi: 10.1088/0031-9155/46/3/201.
- [84] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in *Advances in Neural Information Processing Systems*, 2012, vol. 25, [Online]. Available: <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>.
- [85] Simonyan K and Zisserman A, “Very deep convolutional networks for large-scale image recognition Int,” 2015.
- [86] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [87] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.
- [88] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4\_28.
- [89] P. Baldi, “Autoencoders, Unsupervised Learning, and Deep Architectures,” 2012.
- [90] M. Tschannen, O. Bachem, and M. Lucic, “Recent Advances in Autoencoder-Based Representation Learning,” *CoRR*, vol. abs/1812.0, 2018, [Online]. Available: <http://arxiv.org/abs/1812.05069>.
- [91] G. Litjens *et al.*, “A survey on deep learning in medical image analysis.,” *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.
- [92] R. D. Hjelm, S. M. Plis, and V. D. Calhoun, “Variational Autoencoders for Feature Detection of Magnetic Resonance Imaging Data,” *CoRR*, vol. abs/1603.0, 2016, [Online]. Available: <http://arxiv.org/abs/1603.06624>.
- [93] D. Zimmerer, S. A. A. Kohl, J. Petersen, F. Isensee, and K. H. Maier-hein, “Context-encoding Variational Autoencoder for Unsupervised Anomaly Detection,” pp. 1–13.
- [94] A. Deshpande, J. Lu, M.-C. Yeh, and D. A. Forsyth, “Learning Diverse Image Colorization,” *CoRR*, vol. abs/1612.0, 2016, [Online]. Available: <http://arxiv.org/abs/1612.01958>.

- [95] C. L. Giles, G. M. Kuhn, and R. J. Williams, “Dynamic recurrent neural networks: Theory and applications,” *IEEE Trans. Neural Networks*, vol. 5, no. 2, pp. 153–156, 1994, doi: 10.1109/TNN.1994.8753425.
- [96] J. Chung, Ç. Gülçehre, K. Cho, and Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” *CoRR*, vol. abs/1412.3, 2014, [Online]. Available: <http://arxiv.org/abs/1412.3555>.
- [97] K. Cho, B. van Merriënboer, Ç. Gülçehre, F. Bougares, H. Schwenk, and Y. Bengio, “Learning Phrase Representations using {RNN} Encoder-Decoder for Statistical Machine Translation,” *CoRR*, vol. abs/1406.1, 2014, [Online]. Available: <http://arxiv.org/abs/1406.1078>.
- [98] T. Mansi and R. Liao, “Dilated FCN for Multi-Agent 2D / 3D Medical Image Registration.”
- [99] M. T. and L. Miao S, Piat S, Fischer PW, Tuysuzoglu A, Mewes PW, “Dilated FCN for multi-agent 2D/3D medical image registration.”
- [100] I. J. Goodfellow, J. Pouget-abadie, M. Mirza, B. Xu, and D. Warde-farley, “Generative Adversarial Nets,” pp. 1–9.
- [101] Y. Lu *et al.*, “CT-TEE Image Registration for Surgical Navigation of Congenital Heart Disease Based on a Cycle Adversarial Network.,” *Comput. Math. Methods Med.*, vol. 2020, p. 4942121, 2020, doi: 10.1155/2020/4942121.
- [102] D. Mahapatra, S. Sedai, and R. Garnavi, “Elastic Registration of Medical Images With GANs,” *CoRR*, vol. abs/1805.0, 2018, [Online]. Available: <http://arxiv.org/abs/1805.02369>.
- [103] Y. Lei *et al.*, “CT prostate segmentation based on synthetic MRI-aided deep attention fully convolution network.,” *Med. Phys.*, vol. 47, no. 2, pp. 530–540, Feb. 2020, doi: 10.1002/mp.13933.
- [104] X. Dong *et al.*, “Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network.,” *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol.*, vol. 141, pp. 192–199, Dec. 2019, doi: 10.1016/j.radonc.2019.09.028.
- [105] X. Cheng, L. Zhang, and Y. Zheng, “Deep similarity learning for multimodal medical images,” *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.*, vol. 6, no. 3, pp. 248–252, May 2018, doi: 10.1080/21681163.2015.1135299.
- [106] G. Haskins *et al.*, “Learning Deep Similarity Metric for 3D {MR-TRUS} Registration,” *CoRR*, vol. abs/1806.0, 2018, [Online]. Available: <http://arxiv.org/abs/1806.04548>.
- [107] A. Sedghi, T. Kapur, J. Luo, and P. Mousavi, “Probabilistic Image Registration via Deep Multi-class Classification: Characterizing Uncertainty Probabilistic Image Registration via Deep Multi-class Classification: Characterizing Uncertainty,” no. October, pp. 0–10, 2019, doi: 10.1007/978-3-030-32689-0.
- [108] E. C. And and Z. Wu, “AIRNet: Self-Supervised Affine Registration for 3D Medical Images using Neural Networks,” *CoRR*, vol. abs/1810.0, 2018.
- [109] Z. Yao *et al.*, “A supervised network for fast image-guided radiotherapy (IGRT)

- registration,” *J. Med. Syst.*, vol. 43, no. 7, p. 194, 2019, doi: 10.1007/s10916-019-1256-y.
- [110] H. Guo, M. Kruger, S. Xu, B. J. Wood, and P. Yan, “Computerized Medical Imaging and Graphics Deep adaptive registration of multi-modal prostate images,” *Comput. Med. Imaging Graph.*, vol. 84, no. July, p. 101769, 2020, doi: 10.1016/j.compmedimag.2020.101769.
- [111] S. K. Z. Haofu Liao, Wei-An Lin, Jiarui Zhang, Jingdan Zhang, Jiebo Luo, “Multiview 2D/3D Rigid Registration via a Point-Of-Interest Network for Tracking and Triangulation (POINT2),” *CoRR*, vol. abs/1903.0, 2019, [Online]. Available: <http://arxiv.org/abs/1903.03896>.
- [112] S. S. M. S. And, S. K. And, D. E. And, and A. Gholipour, “Real-time Deep Registration With Geodesic Loss,” *CoRR*, vol. abs/1803.0, 2018.
- [113] H. Guo, X. Xu, S. Xu, B. Wood, and P. Yan, “End-to-end Ultrasound Frame to Volume Registration,” vol. 16, no. (3), pp. 642–61, 2021, doi: 10.1016/j.media.2010.03.005. Epub 2010 Apr 13.
- [114] Y. Fu, Y. Lei, T. Wang, M. Axente, and J. Roper, “Deep learning based volume-to-slice MRI registration via intentional overfitting,” vol. 1203412, no. April, 2022, doi: 10.1117/12.2611898.
- [115] Z. Sandoval, M. Castro, J. Alirezaie, F. Bessière, C. Lafon, and J.-L. Dillenseger, “Transesophageal 2D Ultrasound to 3D Computed Tomography registration for the guidance of a cardiac arrhythmia therapy,” *Phys. Med. Biol.*, vol. 63, no 15, p. 155007, 2018.
- [116] U. K. Lopes and J. F. Valiati, “Pre-trained convolutional neural networks as feature extractors for tuberculosis detection,” *Comput. Biol. Med.*, vol. 89, no. February, pp. 135–143, 2017, doi: 10.1016/j.compbiomed.2017.08.001.
- [117] H. Uzunova, M. Wilms, H. Handels, and J. Ehrhardt, “Training CNNs for Image Registration from Few Samples with Model-based Data Augmentation,” in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, 2017, pp. 223–231.
- [118] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, “End-to-end unsupervised deformable image registration with a convolutional neural network,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10553 LNCS, pp. 204–212, 2017, doi: 10.1007/978-3-319-67558-9\_24.
- [119] J. Lv, M. Yang, J. Zhang, and X. Wang, “Respiratory motion correction for free-breathing 3D abdominal MRI using CNN-based image registration: a feasibility study.,” *Br. J. Radiol.*, vol. 91, no. 1083, p. 20170788, Feb. 2018, doi: 10.1259/bjr.20170788.
- [120] M. Simonovsky, “A Deep Metric for Multimodal Registration,” pp. 1–10.
- [121] Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial Transformer Networks,” *28th Int. Conf. Neural Inf. Process. Syst.*, vol. 2, pp. 2017–2025, 2015.
- [122] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca,



- “VoxelMorph: A Learning Framework for Deformable Medical Image Registration,” *IEEE Trans. Med. Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019, doi: 10.1109/TMI.2019.2897538.
- [123] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, “Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain.,” *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, Feb. 2008, doi: 10.1016/j.media.2007.06.004.
- [124] J. A. Shackleford, N. Kandasamy, and G. C. Sharp, “On developing B-spline registration algorithms for multi-core processors.,” *Phys. Med. Biol.*, vol. 55, no. 21, pp. 6329–6351, Nov. 2010, doi: 10.1088/0031-9155/55/21/001.
- [125] C. Qin, W. Bai, J. Schlemper, S. E. Petersen, K. Stefan, and C. V Jun, “Joint Learning of Motion Estimation and Segmentation for Cardiac MR Image Sequences.”
- [126] J. Zhang, “Inverse-Consistent Deep Networks for Unsupervised Deformable Image Registration,” *CoRR*, vol. abs/1809.0, 2018, [Online]. Available: <http://arxiv.org/abs/1809.03443>.
- [127] D. P. Huttenlocher, W. J. Rucklidge, and G. A. Klanderman, “Comparing images using the Hausdorff distance under translation,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1992-June, pp. 654–656, 1992, doi: 10.1109/CVPR.1992.223209.
- [128] B. Kim, J. Kim, J.-G. Lee, D. H. Kim, S. H. Park, and J. C. Ye, “Unsupervised Deformable Image Registration Using Cycle-Consistent CNN,” *CoRR*, vol. abs/1907.0, 2019, [Online]. Available: <http://arxiv.org/abs/1907.01319>.
- [129] M. Rohé *et al.*, “SVF-Net: Learning Deformable Image Registration Using Shape Matching To cite this version: HAL Id: hal-01557417 SVF-Net: Learning Deformable Image Registration Using Shape Matching,” 2017.
- [130] M. Lorenzi, N. Ayache, G. B. Frisoni, and X. Pennec, “LCC-Demons: a robust and accurate symmetric diffeomorphic registration algorithm.,” *Neuroimage*, vol. 81, pp. 470–483, Nov. 2013, doi: 10.1016/j.neuroimage.2013.04.114.
- [131] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, “Elastix: A toolbox for intensity-based medical image registration,” *IEEE Trans. Med. Imaging*, vol. 29, no. 1, pp. 196–205, 2010, doi: 10.1109/TMI.2009.2035616.
- [132] S. Sun, L. and Zhang, “Deformable mri-ultrasound registration using 3d convolutional neural network.,” *Image Process. Ultrasound Syst. Assist. Diagnosis Navig.*, p. 152{158, 2018.
- [133] A. Hering, S. Kuckertz, S. Heldmann, and M. P. Heinrich, “Memory-efficient 2.5D convolutional transformer networks for multi-modal deformable registration with weak label supervision applied to whole-heart CT and MRI scans.,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 14, no. 11, pp. 1901–1912, Nov. 2019, doi: 10.1007/s11548-019-02068-z.
- [134] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, 2015.
- [135] B. Dahman, “High Intensity Focused Ultrasound Therapy Guidance System by

- Image-based Registration for Patients with Cardiac Fibrillation,” *CinC*, vol. 46, 2019.
- [136] K. Marstal, F. Berendsen, M. Staring, and S. Klein, “SimpleElastix: A User-Friendly, Multi-lingual Library for Medical Image Registration,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 574–582, 2016, doi: 10.1109/CVPRW.2016.78.
- [137] L. R. Dice, “Measures of the Amount of Ecologic Association Between Species,” *Ecology*, vol. 26, no. 3, pp. 297–302, 1945, doi: 10.2307/1932409.



## Annexe I.

### Quaternion:

**Definition.** The quaternion are the extension of complex numbers. Sir William R. Hamilton (1843) try to find a set of numbers that shared in 3-D space properties similar to those of complex numbers in the plane. We will show the application of quaternion to deal with 3-D rotations [de Casteljaou, 1987, Horn, 1987, Simo et al., 1988, Reyes-Avila, 1990, Reyes-Avila, 1991].

The set  $\mathbb{H}$  (equals  $\mathbb{R}^4$ ) is a four-dimensional normed division algebra over the real numbers. Its canonical basis is  $(1, i, j, k)$ . A quaternion is a number  $\lambda = a + ib + jc + kd$ .

The addition of two quaternions  $\lambda_1 = a_1 + ib_1 + jc_1 + kd_1$  and  $\lambda_2 = a_2 + ib_2 + jc_2 + kd_2$  is  $\lambda_1 + \lambda_2 = (a_1 + a_2) + i(b_1 + b_2) + j(c_1 + c_2) + k(d_1 + d_2)$ .  $(\mathbb{H}, +)$  is an additive commutative group.

The multiplication of the basis elements are:

$$1.1 = 1; 1.i = i; 1.j = j; 1.k = k;$$

$$i^2 = j^2 = k^2 = -1;$$

$$i.j = -j.i = k; j.k = -k.j = i; k.i = -i.k = j;$$

The product  $*$  of two quaternions  $\lambda_1 = a_1 + ib_1 + jc_1 + kd_1$  and  $\lambda_2 = a_2 + ib_2 + jc_2 + kd_2$  is:  $\lambda_1 * \lambda_2 = (a_1 + ib_1 + jc_1 + kd_1)(a_2 + ib_2 + jc_2 + kd_2) = a_1a_2 - b_1b_2 - c_1c_2 - d_1d_2 + (a_1b_2 + b_1a_2 + c_1d_2 - d_1c_2)i + (a_1c_2 - b_1d_2 + c_1a_2 + d_1b_2)j + (a_1d_2 + b_1c_2 - c_1b_2 + d_1a_2)k$ .  $(\mathbb{H}, *)$  is a non commutative multiplicative group.

**Vector subspaces.** The set  $\mathbb{H}$  can be written as a set of quadruples:  $\mathbb{H} = \{(a, b, c, d) | a, b, c, d \in \mathbb{R}\}$ . We can define 2 vector subspaces of  $\mathbb{H}$ :

- $\mathbb{H}_R = \{(a, 0, 0, 0) | a \in \mathbb{R}\} \subset \mathbb{H}$ . It corresponds to the real number.  $\mathbb{H}_R$  is an isomorphism to  $\mathbb{R}$ .
- $\mathbb{H}_V = \{(0, b, c, d) | b, c, d \in \mathbb{R}\} \subset \mathbb{H}$ . It corresponds to the pure imaginary number.  $\mathbb{H}_V$  isomorphism to  $\mathbb{R}^3$ .

Any quaternion  $\lambda = a + ib + jc + kd$  can be decomposed into  $a$  (scalar part of  $\lambda$ ) and  $ib + jc + kd$  (vectorial part of  $\lambda$ ).

So  $\lambda = (r, \mathbf{p})$  with  $r \in \mathbb{R}$  the real and  $\mathbf{p} \in \mathbb{R}^3$  the imaginary part. Any vector  $\mathbf{p} \in \mathbb{R}^3$  can be associated with a pure imaginary number  $\lambda = (0, \mathbf{p})$ .

The product of two quaternions  $\lambda_1 = (r_1, \mathbf{p}_1)$  and  $\lambda_2 = (r_2, \mathbf{p}_2)$  is  $\lambda_1 * \lambda_2 = (r_1, \mathbf{p}_1)(r_2, \mathbf{p}_2) = (r_1r_2 - \mathbf{p}_1 \cdot \mathbf{p}_2, \mathbf{p}_1 \times \mathbf{p}_2 + r_1\mathbf{p}_2 + r_2\mathbf{p}_1)$  with  $\cdot$  and  $\times$  respectively the dot product and the cross product in  $\mathbb{R}^3$ .

- Conjugate:  $\lambda = (r, \mathbf{p}) \rightarrow \bar{\lambda} = (r, -\mathbf{p})$
- Norm:  $\|\lambda\|^2 = \lambda * \bar{\lambda} = \bar{\lambda} * \lambda = a^2 + \|\mathbf{p}\|^2 = a^2 + b^2 + c^2 + d^2$
- Inverse  $\lambda^{-1} = \frac{\bar{\lambda}}{\|\lambda\|^2}$
- Unitary quaternion. It is defined by  $\|\lambda\|^2 = 1$ . It can be deduced that unitary quaternion  $\lambda^{-1} = \bar{\lambda}$ . The group of unitary quaternions forms a sphere  $S^3$  of dimension 3.

**Parametric rotation representation.** Let define the following linear transformation  $R$  of a fixed  $p$ :

$$R(p, ): \mathbb{H} \times \mathbb{H} \rightarrow \mathbb{H}$$

$$R(p, q) = p * q * p^{-1} = \frac{1}{\|p\|^2} (p * q * \bar{p})$$

$R$  is a linear and orthonormal transformation. If  $q \in \mathbb{H}_V$  then  $R(p, q) \in \mathbb{H}_V$ .

For  $\lambda$  an unitary quaternion and  $q \in \mathbb{H}_V$ :

$$R(\lambda, ): S^3 \times \mathbb{H}_V \rightarrow \mathbb{H}_V$$

$$R(\lambda, q) = \lambda * q * \lambda^{-1} = \lambda * q * \bar{\lambda}$$

This restriction of  $R$  in  $\mathbb{R}^3$  is a rotation  $R$  in  $\mathbb{R}^3$  in which all the geometric characteristics (rotation axis  $\mathbf{n}$  and rotation angle  $\theta$ ) are represented by:

$$\lambda = (\cos(\theta/2), \mathbf{n} \sin(\theta/2))$$

**Quaternion and rotation matrix.** The transformation between a quaternion representative of a rotation and a rotation matrix are:

- Quaternion to rotation matrix. For  $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$

$$R = \begin{bmatrix} \lambda_0^2 + \lambda_1^2 - \lambda_2^2 - \lambda_3^2 & 2(\lambda_1\lambda_2 - \lambda_0\lambda_3) & 2(\lambda_1\lambda_3 + \lambda_0\lambda_2) & 0 \\ 2(\lambda_1\lambda_2 + \lambda_0\lambda_3) & \lambda_0^2 + \lambda_2^2 - \lambda_1^2 - \lambda_3^2 & 2(\lambda_2\lambda_3 - \lambda_0\lambda_1) & 0 \\ 2(\lambda_1\lambda_3 - \lambda_0\lambda_2) & 2(\lambda_2\lambda_3 + \lambda_0\lambda_1) & \lambda_0^2 + \lambda_3^2 - \lambda_1^2 - \lambda_2^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- Rotation matrix to quaternion.  $R$  a  $3 \times 3$  rotation matrix (unitary and positive).

$$tr(R) = 1 + 2\cos\theta \text{ (tr: trace)}$$

$$\frac{1}{2}(R - R^T) = \sin\theta \cdot \mathbf{n}$$

$$\lambda_0 = \frac{1}{2}\sqrt{1 + tr(R)}$$

$$\lambda_1 = \frac{1}{4\lambda_0}(R_{32} - R_{23})$$

$$\lambda_2 = \frac{1}{4\lambda_0}(R_{13} - R_{31})$$

$$\lambda_3 = \frac{1}{4\lambda_0}(R_{21} - R_{12})$$



---

**Titer:** Guidage par l'image d'une sonde HIFU transoesophagienne pour le traitement des arythmies cardiaques

**Mots clés :** guidage de thérapies par l'image, recalage multimodal d'images, ablation par ultrasons focalisés de haute intensité, arythmie cardiaque.

**Résumé :** Les ultrasons focalisés de haute intensité (HIFU) par voie transoesophagienne peuvent être utilisés pour traiter l'arythmie cardiaque de manière efficace et non invasive. L'œsophage étant situé juste aux abords du cœur, il offre une fenêtre acoustique parfaite pour que les HIFU puissent être dirigés vers le cœur afin de mener l'ablation nécessaire au traitement de l'arythmie. Cependant, afin de guider l'ablation, le thérapeute doit faire le lien, par recalage, entre les images per-opératoires (images échographiques fournies par la sonde HIFU dual-mode) et l'imagerie anatomique préopératoire à haute résolution, dans laquelle la ligne d'ablation a été définie (volume scanner X ou IRM). Dans ce travail de thèse, nous avons proposé plusieurs solutions pour améliorer ce recalage pendant la chirurgie si possible en temps réel. Premièrement, nous avons intégré un second plan image perpendiculaire au premier

dans la solution classique itérative de recalage. Ensuite, nous nous sommes concentrés sur le recalage rigide d'une image échographique 2D dans un volume scanner X 3D à l'aide d'une approche par apprentissage profond supervisé afin d'estimer la position en temps réel de la sonde d'imagerie/thérapie pendant la chirurgie. L'utilisation d'un réseau a permis d'effectuer le recalage sur des paires d'images inconnues de manière non itérative réduisant ainsi drastiquement le temps de calcul de la méthode classique. En dernier, nous avons abordé une approche d'apprentissage non supervisée. Cette étude visait à effectuer un recalage non rigide entre l'image échographique et une coupe du scanner X afin de prendre en compte, pour une certaine phase du cycle cardiaque, les légères déformations du cœur résultant de la respiration du patient ou de l'insertion de la sonde.

---

**Title:** Image-based guidance of a transesophageal HIFU probe for the treatment of cardiac arrhythmias

**Keywords:** Image-guidance therapy, multimodal image registration, high intensity focused ultrasound ablation, cardiac arrhythmia.

**Abstract:** Transesophageal high intensity focused ultrasound (HIFU) can be used to treat cardiac arrhythmias efficiently and non-invasively. Since the esophagus is located right next to the heart, it provides a perfect acoustic window for HIFU to be directed toward the heart to perform the ablation necessary to treat the arrhythmia. However, in order to guide the ablation, the therapist has to make the link, by registration, between the intraoperative images (ultrasound – US – images provided by the dual-mode HIFU probe) and the preoperative high-resolution anatomical imaging, in which the ablation line has been defined (CT or MRI volumes). In this thesis work, we proposed several solutions to improve the image guidance during the real time surgery. First, we present an iterative framework to estimate the positioning of a new dual probe

in the preoperative 3D CT scan. Second, we focused on rigid registration of a 2D ultrasound image in a 3D X-ray volume using a supervised deep learning approach to estimate the real-time position of the imaging/therapy probe during the surgery. The use of a network allowed us to perform the registration on unknown image pairs in a non-iterative way thus drastically reducing the computational time compared to the classical method. And finally, we developed an unsupervised learning approach. This study aimed at performing a non-rigid registration between 2D US and 2D CT images in order to take into account, for a certain phase of the cardiac cycle, the slight deformations of the heart resulting from the breathing of the patient or the insertion of the probe.