Real-time tracking of deformable targets in 3D ultrasound sequences

Lucas Royer

Thèse

THESE INSA Rennes
sous le sceau de l'Université Bretagne Loire
pour obtenir le titre de
DOCTEUR DE L'INSA RENNES
Spécialité : Traitement du signal et de l'image

présentée par
Lucas Royer

ECOLE DOCTORALE : *MATISSE*
LABORATOIRE : *IRISA*

Real-time Tracking of Deformable Targets in 3D Ultrasound Sequences

Thèse soutenue le 06.12.2016
devant le jury composé de :

Stéphane Cotin
Directeur de Recherche, Inria, Strasbourg/ Président
Jocelyne Troccaz
Directrice de Recherche, TIMC-CNRS, Grenoble/ Rapporteur
Hervé Liebgott
Professeur, Université Claude Bernard- CREATIS, Lyon/ Rapporteur
Guillaume Dardenne
Docteur, CHU, Brest / Co-Encadrant de thèse
Maud Marchal
Maitre de Conférences, IRISA-INSA, Rennes/ Co-Encadrante de thèse
Alexandre Krupa
Chargé de Recherche, Inria, Rennes/ Directeur de thèse

Résumé en Français

Introduction

Cette thèse, intitulée «suivi temps-réel de cibles dans des séquences de volumes échographiques», propose plusieurs contributions permettant d'estimer la position de structures anatomiques dans des images échographiques 3D. Ces contributions sont liées à plusieurs domaines de recherches tels que la vision assistée par ordinateur, la simulation physique et le traitement d'image médicale. Cette thèse a été réalisée au sein de l'IRT b<>com et a impliqué la collaboration du centre hospitalier universitaire de Rennes, et des centres de recherche Irisa et Inria Rennes-Bretagne Atlantique. L'application visée consiste à améliorer la visualisation permettant de guider des interventions mini-invasives dédiées au traitement du cancer du foie.

Avec près de 800.000 décès par an, le cancer du foie est le sixième le plus mortel dans le monde. Au cours des dernières années, la prise en charge de ce cancer a évolué et plusieurs stratégies de traitements sont aujourd'hui envisageables telles que l'ablation partielle et la greffe. De nos jours, les interventions mini-invasives sont de plus en plus utilisées, car elles permettent un traitement plus local. Ainsi, ces procédures ont l'avantage de réduire le temps de rémission des patients et les risques de complication. Parmi ces traitements, l'ablation par radio-fréquence, la cryothérapie et l'injection percutanée d'éthanol sont les interventions les plus courantes. Ces techniques sont généralement basées sur l'insertion d'une aiguille permettant la destruction thermique ou chimique des tissus cancéreux. Afin d'assurer un placement correct de l'aiguille par rapport aux tissus, ces interventions sont généralement guidées par différents types d'imagerie tels que le scanner, l'imagerie par résonance magnétique, la fluoroscopie, ou l'échographie. Cependant, l'imagerie échographique est souvent préférée car cette modalité est non-invasive, bas coût et portable. De plus, elle permet d'obtenir un retour visuel temps-réel des tissus au cours de l'intervention. Elle est donc très adaptée pour des applications cliniques nécessitant de visualiser la cible clinique et les instruments médicaux tout au long de l'intervention.

Malgré le fort développement des traitements mini-invasifs, leur efficacité est toujours limitée par plusieurs problèmes liés à la précision de l'intervention. En effet, contrairement à la chirurgie classique, ces thérapies ne peuvent délivrer qu'un traitement local autour de la pointe de l'aiguille. Ainsi, un mauvais placement de l'aiguille par rapport aux tissus peut compromettre le succès de l'intervention. Un autre problème est lié aux déplacements et aux déformations des tissus liés aux mouvements physiologiques du patient et à la manipulation d'instruments médicaux. Enfin, malgré les différents avantages de l'imagerie échographique, cette modalité présente plusieurs limitations compliquant la visualisation de l'aiguille et des structures cliniques visées. En effet, l'imagerie échographique présente un champ de vue limité et est généralement perturbée par plusieurs artefacts tels que le bruit de type «speckle », les effets d'ombres et de réverbérations. Afin de pallier ces problèmes, de nombreuses équipes de recherches ont proposé des travaux permettant d'estimer la position de structures anatomiques dans les images échographiques. Ce travail vise à proposer de nouvelles contributions permettant une estimation plus précise et plus robuste des structures d'intérêts.

Etat de l'art

Dans cette section, nous présentons différentes approches permettant de suivre des régions d'intérêts dans une séquence d'images échographiques. Ces méthodes ont pour objectif de déterminer successivement la transformation permettant d'aligner une image de référence et une image courante. Elles reposent généralement sur l'optimisation d'une fonction de coût dépendante de plusieurs caractéristiques. Dans la suite de cette section, nous présentons les contributions principales de ce domaine selon trois caractéristiques: le modèle de transformation, le critère de similarité et le terme de régularisation.

Modèle de transformation

Selon les tissus observés par l'imagerie échographique, une structure d'intérêt peut subir différents types de mouvements allant du simple déplacement rigide à la déformation. En conséquence, plusieurs modèles de transformation ont été étudiés afin de représenter l'espace des déplacements possibles. Parmi les plus simples, le modèle translationnel (Veronesi et al., 2005), le modèle rigide (Rothlubbers et al., 2014) et le modèle affine (Wein et al., 2008b) sont souvent utilisés. Cependant, on peut noter que ces modèles de transformation ne sont pas adaptés au suivi de structures déformables. Pour pallier ce problème, des modèles de transformation plus sophistiqués ont été utilisés. Par exemple, certaines approches reposent

sur le modèle de type «block-matching »permettant d'estimer le déplacement d'une cible en décomposant son mouvement en plusieurs blocs rigides (Basarab et al., 2008; De Luca et al., 2013; Touil et al., 2010; Yeung et al., 1998). En revanche, ce type de transformation ne permet pas de représenter des déformations très localisées puisqu'il considère que le mouvement est constant à l'intérieur d'un bloc. Pour résoudre ce problème, des modèles déformables ont été proposés tels que les transformations de type plaque mince (Lee and Krupa, 2011) et les modèles de déformation de forme libre (Heyde et al., 2012; Ledesma-Carbayo et al., 2001; Pennec et al., 2001). Alors que ces transformations permettent de représenter correctement différents types de mouvement, elles sont généralement associées à un grand nombre de paramètres ce qui les rend particulièrement sensible aux problèmes de minima locaux. Certaines méthodes de suivi sont basées uniquement sur des méthodes de segmentation permettant de définir un contour autour de la région d'intérêt. Ces contours sont généralement représentés par des fonctions implicites (Dietenbeck et al., 2014) ou explicites (Nascimento and Marques, 2008).

Critère de similarité

Le critère de similarité permet de mesurer l'alignement entre une image de référence et une image courante transformée. L'optimisation de ce critère a pour but de définir les paramètres optimaux de la transformation alignant deux images. Généralement, les mesures de similarité peuvent être décomposées en deux catégories. Le premier type évalue la distance entre différentes caractéristiques extraites dans les images telles que des points d'intérêts (Schneider et al., 2012), des contours 2D (Angelova and Mihaylova, 2010), des surfaces 3D (Papademetris et al., 2002). Pour mesurer cette distance, la plupart des approches utilise la norme Euclidienne (Häme et al., 2012). Cependant, d'autres types de mesure ont aussi été étudiés tel que la distance Mahalanobis permettant d'être plus robuste aux valeurs aberrantes (Comaniciu et al., 2004). La deuxième catégorie permet d'évaluer l'alignement entre deux images à partir de la mesure de correspondance d'intensité. La plupart de ces approches considèrent des critères monomodales tel que la Somme du Carrée des Différences (SSD) (Lubke and Grozea, 2014; Royer et al., 2015; Yeung et al., 1998), la somme des Différences Absolues (SAD) (Touil et al., 2010), ou la Corrélation Croisée (CC) (Basarab et al., 2008; De Luca et al., 2013). Ces critères ont l'avantage d'être très robustes lorsque l'intensité de la cible reste constante au cours du temps. En revanche, des variations d'intensités peuvent être observées durant les interventions à cause de plusieurs artefacts liés à l'imagerie échographique. Afin de pallier ce problème, différentes approches se sont basées sur des mesures de similarité spécifiques à l'imagerie échographique (Baumann et al., 2012; Cohen and Dinstein, 2002). Ces dernières ont l'avantage d'être très robustes aux variations

d'intensité globales et au bruit de type «speckle »affectant les images échographiques. En revanche, ces mesures n'ont pas été validées sur des séquences contenant de fortes variations d'intensité locales introduites par la présence d'ombres dans l'imagerie ultrasonore. Afin de contourner ce problème, différentes approches se sont appuyées sur l'utilisation de critères multimodaux tel que l'information mutuelle (Elen et al., 2008; Shekhar and Zagrodsky, 2002) et la somme de variance conditionnelle (Masum et al., 2014). Malgré la performance du critère d'information mutuelle, cette mesure n'est pas adaptée à des applications de suivi temps-réel déformable car elle requiert un fort temps de calcul. En revanche, la somme de variance conditionnelle a l'avantage d'être rapide, mais sa performance est impactée lorsque des variations d'intensité locale sont prises en compte.

Terme de régularisation

Lorsqu'un critère de similarité est bien défini, il doit en principe permettre d'obtenir une transformation optimale alignant l'image de référence et l'image courante. Cependant, plusieurs problèmes, liés aux bruits et aux minima locaux, peuvent conduire à la mauvaise estimation de cette transformation. Afin de pallier ces limitations, différentes approches utilisent un terme de régularisation permettant de pénaliser les transformations non réalistes suivant des mouvements non-continus, des repliements, ou des étirements ou rétrécissements brutaux. Ces régularisations peuvent être basées sur des contraintes spatio-temporelles permettant d'assurer le lissage du champ de déplacements (Mukherjee et al., 2011). D'autres méthodes utilisent une contrainte de cohérence d'inversion permettant d'assurer que la composition de la transformation par son inverse est proche de l'identité (Baumann et al., 2012; Vijayan et al., 2013). D'autres méthodes permettent d'assurer que la région d'intérêt ne subit pas de variation de volume au cours du temps (Elen et al., 2008). Afin de fournir des résultats robustes au cours de l'acquisition, certaines méthodes sont basées sur la prédiction temporelle grâce à l'utilisation de filtre de Kalman (Orderud et al., 2007), ou de filtre à particules (Angelova and Mihaylova, 2010). La prédiction temporelle est particulièrement adaptée à la détection de caractéristiques, car ces dernières ne sont pas toujours visibles au cours du temps. Enfin, quelques approches proposent une régularisation basée sur une détection de champs de déplacement aberrants. Par exemple, Banerjee et al. (2015) propose une méthode permettant d'inhiber certains mouvements selon des contraintes de géométrie et d'apparence.

Limitations et verrous technologiques

Malgré le nombre de contributions proposées dans ce domaine, il est toujours difficile de déterminer l'approche permettant d'obtenir les meilleurs résultats pour une application précise. En effet, ces approches sont généralement évaluées sur des quantités de données inégales obtenues à partir de systèmes d'imagerie différents. De plus, certaines ne sont pas évaluées sur données réelles humaines. Un autre problème est lié à la complexité entre les différentes bases de données d'images échographique utilisées. En effet, nous avons remarqué que très peu d'approches sont validées sur des images d'évaluations affectées par différents artefacts de l'imagerie échographique tel que les zones d'ombres et les effets miroirs. Afin de pallier ces problèmes, nous proposons plusieurs contributions permettant d'estimer de façon robuste la position de structures d'intérêts au cours du temps. Elles incluent un nouveau critère de similarité, une nouvelle stratégie de suivi, ainsi qu'une méthode basée sur la simulation physique. La performance de ces contributions est évaluée sur différentes bases de données acquises par simulation, sur maquettes simulant des tissus mous et sur volontaires. De plus, nous avons comparé notre approche par rapport à certaines méthodes de l'état de l'art en évaluant sa précision sur différentes bases de données 3D fournies par les challenges MICCAI CLUST 2014 et 2015.

Approche de suivi déformable basée sur modèle physique

Dans cette section, nous proposons une approche temps-réel permettant d'estimer la position de structures d'intérêts dans une séquence de volumes échographiques. L'approche repose sur l'utilisation conjointe de l'information visuelle dense (intensité des pixels de l'image) et d'une méthode de simulation physique. Afin d'évaluer la performance de cette méthode, différentes expériences ont été réalisées sur données acquises sur maquettes simulant des tissus mous et sur volontaires humains.

Description de la méthode

L'approche proposée permet de calculer les déplacements d'un modèle représentant la structure d'intérêt. Ces déplacements sont obtenus en sommant itérativement des déplacements internes estimés à partir d'une approche mécanique et des déplacements externes calculés grâce à une approche utilisant l'intensité des pixels/voxels de l'image. Cette méthode peut être décomposée en trois étapes:

- **Génération du modèle:** Dans un premier temps, un modèle représentant la structure d'intérêt est généré dans le premier volume de la séquence échographique. Pour ce

faire, nous réalisons une segmentation manuelle délimitant la surface de la cible sur chaque coupe du volume échographique. A partir de cette segmentation, nous générons un modèle 3D composé de cellules tétraédriques et de sommets. Ce modèle permet de définir la position de chaque voxel par rapport à la position des sommets des tétraèdres.

- **Estimation des déplacements externes:** Une fois que le modèle initial est obtenu, les déplacements de ses sommets sont estimés successivement au cours du temps. Pour cela, l'approche calcule premièrement les déplacements externes à partir d'une méthode permettant de minimiser la différence d'intensité entre une image courante et une image de référence. À cette fin, une fonction de coût basée sur le critère SSD est minimisée grâce à l'algorithme de la plus forte pente.

- **Estimation des déplacements internes:** Dû à la présence de certains minima locaux, une mauvaise estimation des déplacements externes peut entrainer une transformation non réaliste. Afin de pallier ce problème, notre méthode estime les déplacements internes à partir de la simulation d'un modèle masse-ressort-amortisseur lié au modèle 3D décrit précédemment. A partir de ce modèle mécanique, nous calculons les forces internes appliquées sur chaque sommet du modèle. Ensuite, les déplacements internes sont finalement obtenus par l'intégration des forces grâce à un schéma d'intégration d'Euler.

Résultats

Grâce à une implémentation sur technologie GPGPU, le temps de calcul de notre méthode atteint 350 ms pour l'obtention de la transformation entre deux images. Afin de valider notre approche, deux types de résultats ont été générés. Les premières expériences, réalisées sur données acquises sur une maquette simulant des tissus mous, ont permis d'évaluer notre approche sur différents types de mouvements. Ces expériences ont été réalisées grâce à un système robotique permettant de déplacer la sonde à la surface d'une maquette contenant un objet organique déformable. La vérité terrain est obtenue grâce à l'odométrie permettant de mesurer les déplacements du robot. Sur la figure 1, nous avons illustré l'estimation de la trajectoire de la cible durant l'une des expériences réalisées. Nous pouvons observer que seule l'approche combinant la méthode de simulation physique permet de suivre correctement la trajectoire définie par l'odométrie.

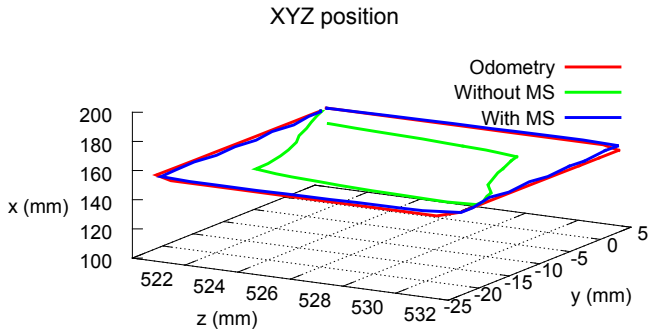


Fig. 1 Estimation de la position de la cible. (Odometry) Vérité terrain. (with MS) Approche combinée à la simulation physique (without MS) Approche non combinée.

Dans un second temps, nous avons comparé notre approche à différentes méthodes de l'état de l'art sur données réelles. Pour réaliser ces expériences, nous avons utilisé les bases de données réelles fournies par les challenges MICCAI CLUST 2014 et 2015. Ainsi, notre approche a été évaluée sur le suivi de 34 structures anatomiques différentes acquises à partir de 22 séquences de volumes échographiques. Les résultats de précision sont présentés dans le tableau 1.

Participants	**Moy**	**ET**	**95%**	**Challenge**
Our method	1.62	2.19	4.81	CLUST 2014
Somphone et al. (2014)	2.55	2.46	7.98	
Rothlubbers et al. (2014)	2.80	2.96	7.94	
Lubke and Grozea (2014)	4.63	4.03	12.44	
Our method	1.74	0.92	3.65	CLUST 2015
Banerjee et al. (2015)	1.80	1.64	3.41	

Table 1 Résultats de l'erreur de suivi exprimés en millimètre sur les bases de données CLUST 2014 et 2015. (Moy) Erreur de suivi moyenne. (ET) Ecart-type. (95%) 95eme centile de l'erreur de suivi.

Critère de similarité

A partir des résultats de la section précédente, nous avons démontré que l'approche proposée est performante et propose un suivi précis sur différents types de structures. Les expériences, réalisées sur maquettes et volontaires, ont permis de prouver que l'approche était robuste au bruit lié à l'acquisition. Cependant, cette méthode n'a pas été évaluée sur des séquences comprenant d'autres artefacts tels que des ombres et des changements de gains dans l'imagerie ultrasonore. Ces derniers peuvent fortement perturber la performance de l'approche car ils introduisent de fortes variations d'intensité dans l'image. Dans cette section, nous proposons un nouveau critère de similarité ainsi qu'une nouvelle stratégie pour résoudre les problèmes précédents. La pertinence de ces deux contributions est évaluée sur données simulées et données acquises sur maquette.

Description de la méthode

Afin d'améliorer la robustesse de notre approche, nous proposons un critère spécifique à l'imagerie échographique basé sur la mesure de similarité dénommée "Somme de Variance Conditionnelle". Cette dernière, proposée par Pickering et al. (2009), mesure la différence d'intensité entre une image courante et une image de référence adaptée. Elle a l'avantage d'être robuste aux variations d'intensité globale affectant la séquence d'image au cours du temps. L'image de référence adaptée est générée à partir d'un opérateur d'espérance qui compense les changements d'intensité entre l'image courante et l'image de référence originale. Différentes expérimentations, proposées par Richa et al. (2011) et Richa et al. (2014), ont permis de démontrer la performance de ce critère par rapport à d'autres mesures de similarité telles que la SSD, ou l'information mutuelle. De plus, la performance de cette mesure de similarité sur l'imagerie échographique a été montrée par Masum et al. (2014). Malgré ses performances, ce critère reste sensible aux ombres générées par ce type d'imagerie. En effet, ce type d'artefact peut créer de forts changements d'intensité locaux générés par la réflexion de l'onde échographique à travers les tissus. Pour pallier ce problème, nous proposons de modifier la mesure de similarité précédente afin d'y inclure une matrice de pondération calculée à partir d'une étape de détection d'ombre. Cette matrice de pondération est introduite dans le calcul de:

- **la fonction de coût** permettant de mesurer la différence d'intensité entre l'image de référence adaptée et l'image courante. L'introduction de la matrice de pondération permet de limiter l'effet des voxels présents dans des zones ombrées en leur associant un poids de pondération faible.

- **l'opérateur d'espérance** basé sur le calcul de la fonction de densité de probabilité jointe entre l'intensité de l'image de référence et l'image courante. L'introduction de cette matrice de pondération dans cette fonction permet de limiter la perturbation des voxels présents dans des zones d'ombres lors de la génération de l'image de référence adaptée.

Résultats

Afin d'évaluer la performance de notre critère, nous proposons de le tester sur des séquences contenant des images fortement perturbées par des variations d'intensité et l'introduction de zones d'ombres. Pour cela, nous utilisons premièrement des données acquises sur maquettes contenant des artefacts synthétiques. La génération de ces artefacts est obtenue en modifiant l'intensité des voxels au cours de la séquence afin de simuler des ombres et des changements de gains. Dans un deuxième temps, nous proposons de valider notre approche sur des séquences réelles acquises sur des maquettes contenant des tissus mous. Les changements de gains et les ombres sont générés par l'utilisateur durant l'acquisition. Dans ces différentes expériences, nous proposons de comparer la précision de notre critère, intitulé somme des variances conditionnelles de confiance (SCCV) par rapport à d'autres mesures de similarité telles que la somme des carrés des différences (SSD), la somme pondérée des carrés des différences (WSSD) et la somme des variances conditionnelles (SCV). Dans les figures et le tableau suivants, nous illustrons quelques résultats de précision obtenus lors de différentes expériences réalisées. Nous pouvons observer que seul notre critère a l'avantage d'être précis et robuste car son erreur reste faible sur toutes les séquences.

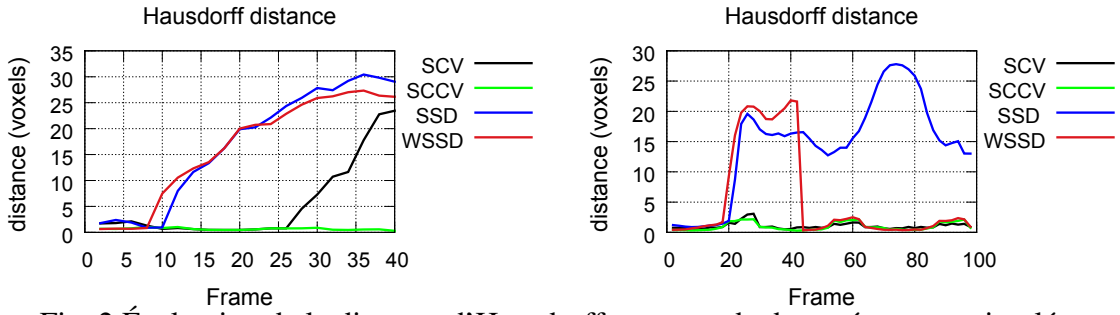


Fig. 2 Évaluation de la distance d'Hausdorff au cours de deux séquences simulées.

Sequence	SSD	WSSD	SCV	SCCV
PHA_1	10.6 ± 11.7	5.8 ± 6.8	22.9 ± 31.1	2.48 ± 2.18
PHA_2	-	5.36 ± 6.01	–	2.0 ± 1.7
PHA_4	–	–	31 ± 41	2.4 ± 2.2

Table 2 Évaluation de l'erreur moyenne de suivi sur trois séquences acquises sur maquette.

Stratégie hybride

Dans la section précédente, nous avons proposé un critère robuste aux variations d'intensité locales et globales générées lors de l'acquisition d'images échographiques. Cette mesure de similarité permet de mesurer la différence d'intensité entre une image de référence et une image courante acquise au cours du temps. En revanche, le choix de l'image de référence peut perturber la performance de l'approche. Pour résoudre ce problème, différentes stratégies ont été étudiées:

- **Les stratégies par paire** sont basées sur l'optimisation d'une fonction de coût classique comparant successivement des paires d'images. Ces approches peuvent être subdivisées en deux types. Un premier type d'approche permet de comparer une image de référence fixe à une image courante acquise au cours du temps. Ce type d'approche est donc sensible au choix de l'image de référence car cette dernière peut induire de mauvais résultats si sa qualité n'est pas suffisante. Un deuxième type d'approche permet de comparer l'image courante à son image précédente ce qui peut provoquer une accumulation de l'erreur au cours du temps.

- **Les stratégies par groupe** sont basées sur l'optimisation d'une fonction de coût comparant toutes les images de la séquence. Ainsi, à chaque nouvelle image acquise, ces approches prennent en compte toute la séquence afin de fournir un résultat de suivi plus robuste. La principale limitation de ces stratégies est liée à la complexité. Ainsi, ce type de stratégie n'est pas adapté pour des applications temps-réel car elle requiert une quantité importante de temps de calcul et de mémoire.

Description de la méthode

Afin de pallier les problèmes des stratégies décrites ci-dessus, nous proposons une approche hybride basée sur les stratégies par paire. Dans cette approche hybride, nous sélectionnons une image de référence correspondant à l'image initiale de la séquence. Mais contrairement aux approches existantes, la méthode proposée permet de remplacer uniquement les zones

d'ombres dans l'image de référence. Ainsi, cette technique assure que la qualité de l'image de référence est maximale au cours du temps. La méthode proposée a l'avantage d'être plus rapide que les méthodes de suivi par groupe. De plus, cette méthode n'est pas sensible à l'accumulation de l'erreur puisqu'elle ne remplace uniquement que les structures de l'image dont la qualité n'est pas suffisante.

Résultats

Afin d'évaluer cette approche, nous avons utilisé des données simulées contenant des artefacts synthétiques, ainsi que des données acquises sur maquette. Cette validation nous permet de comparer les résultats de précision entre différentes stratégies par paire. Dans les figures suivantes, nous illustrons quelques résultats obtenus lors de la validation. Ces résultats montrent l'erreur de suivi entre les différentes stratégies. Nous pouvons observer que seule notre nouvelle stratégie de suivi permet d'obtenir une évolution de l'erreur faible. De plus, son temps de calcul reste constant contrairement aux stratégies par groupes.

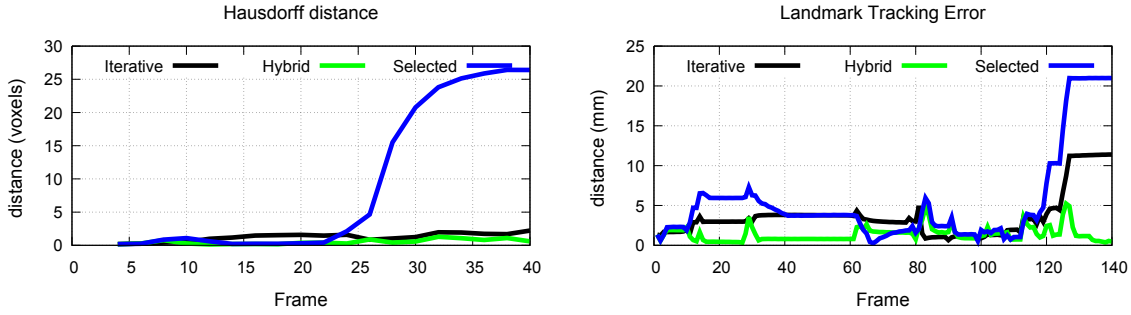


Fig. 3 Évaluation de l'erreur de suivi sur deux séquences. (Selected) Stratégie par paire où l'image de référence est fixe au cours du temps. (Iterative) Stratégie par paire où l'image de référence représente l'image précédente. (Hybrid) Stratégie proposée.

Application pour la visualisation multi-modale

Grâce aux contributions présentées dans les sections précédentes, nous avons proposé une approche permettant de suivre des régions d'intérêts dans des séquences de volumes échographiques. Cette méthode a l'avantage d'être robuste à différents artefacts liés à l'imagerie échographique. Dans cette section, nous présentons une application permettant d'améliorer la visualisation à partir de la fusion de l'imagerie échographique et de l'imagerie par résonance magnétique (IRM). Cette application a été évaluée dans un contexte clinique sur une plateforme liée au Centre Hospitalier Universitaire de Rennes.

Acquisition

Nous avons réalisé consécutivement l'acquisition d'une image IRM et d'une séquence de volumes échographiques sur un volontaire sain. Les volumes échographiques ont été obtenus à partir d'une station Ultrasonix connectée à une sonde 3D motorisée. Le volume IRM a été acquis à partir d'un système d'imagerie 3T Siemens Verio. De plus, un système de localisation, permettant de définir la position du patient par rapport aux repères liés aux différents types d'imagerie, a été utilisé.

Description de la méthode

Dans cette section, nous proposons un système permettant d'améliorer la visualisation échographique en remplaçant les zones ombrées de l'image par des structures visibles dans l'imagerie pré-opératoire. Notre méthode peut être décomposée en trois étapes:

- **Recalage multi-modal**: cette étape consiste à trouver la transformation permettant d'aligner le volume IRM et le volume échographique initial de la séquence. Cette transformation est obtenue à partir d'un recalage externe basé sur un système de localisation optique. Ce dernier permet de relier la position du patient par rapport à la position des systèmes d'imagerie. La transformation est ensuite affinée en utilisant un recalage manuel basé sur l'annotation de points de repères dans les images.

- **Suivi de région d'intérêt**: A partir de l'étape précédente, la transformation permettant d'aligner le premier volume échographique et l'image IRM est obtenue. La seconde étape de la méthode permet de définir une région d'intérêt à partir d'une segmentation dans l'image IRM. Ensuite, nous utilisons l'approche décrite précédemment afin d'estimer la position de la région d'intérêt au cours de la séquence échographique.

- **Visualisation Améliorée**: La dernière étape permet d'améliorer la visualisation en combinant l'imagerie échographique et IRM. Pour ce faire, les régions ombrées de l'image sont premièrement identifiées dans l'imagerie échographique à partir de l'étape de détection d'ombres. Ces zones correspondent aux régions de l'image affectées par des ombres occultant les structures anatomiques. Ensuite, l'intensité de ces régions est remplacée par l'intensité des structures visibles dans l'imagerie IRM.

Résultats

Afin de réaliser ce système, le calibrage des différents systèmes d'imagerie est requis. Ainsi, le système échographique est calibré en adaptant la méthode dénommée N-wire Lasso et al.

(2014). Le calibrage du système IRM est basé sur la détection de marqueurs déposés sur la peau du patient qui sont à la fois visibles dans l'IRM et palpés à partir d'un système de localisation externe. Grâce au système de localisation, un recalage initial approximatif est défini. En revanche, ce type de recalage génère une erreur importante induite par plusieurs facteurs incluant le mouvement respiratoire et la pression de la sonde sur le patient. De tels effets ne peuvent être compensés par le recalage externe lié au localisateur optique. Afin de pallier ce problème, le recalage manuel permet de réduire l'erreur en dessous de 2 mm et permet d'obtenir une bonne initialisation du système de visualisation.

Une fois que le recalage multi-modal est obtenu, nous proposons d'améliorer la visualisation d'une région d'intérêt le long d'une séquence échographique. Afin de fournir la vérité terrain, la position d'un point de repère représentant la bifurcation d'une veine hépatique est annotée sur chaque volume par un expert. La précision des différents critères de similarité est évaluée durant cette tache de suivi. Grâce à l'approche de recalage et de suivi basée sur les contributions précédentes, nous pouvons mettre en place une approche permettant d'améliorer la visualisation en remplaçant les structures ombrées dans l'image échographique au cours du temps. Un exemple de visualisation au cours du temps est illustré sur la figure suivante.

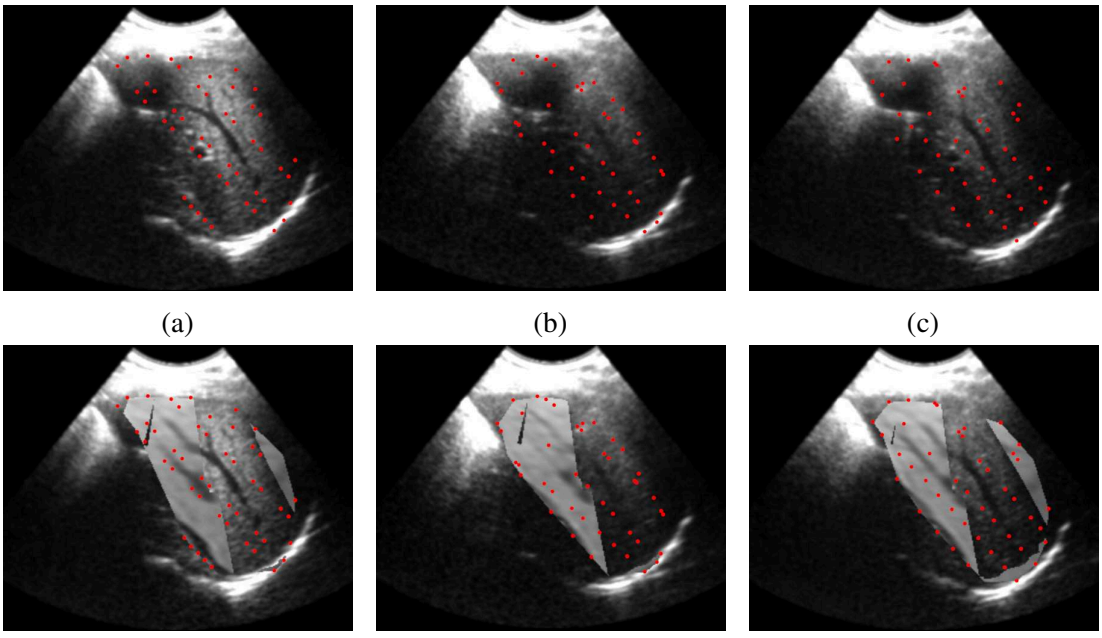


Fig. 4 Exemple du système de visualisation. (a-b-c) Ces images représentent les coupes d'un volume échographique original à différent temps. (d-e-f) Ces images représentent les coupes d'un volume combinant l'imagerie échographique et IRM aux temps associés.

Conclusion

Avec le développement des technologies médicales, les thérapies permettant de traiter le cancer du foie ont considérablement évolué au cours des dernières années. Parmi ces dernières, les interventions mini-invasives sont considérées comme les plus prometteuses car elles permettent un traitement plus localisé. Cependant, leurs succès est toujours impacté par des limitations réduisant la précision de ces interventions. Pour contourner ces problèmes, de nombreuses méthodes, permettant d'estimer la position de structures sous imagerie échographique, ont été étudiées. Dans cette thèse, nous avons proposé une approche reposant sur plusieurs contributions robustes à différents artefacts de l'imagerie échographique. La première contribution combine une approche utilisant conjointement l'information visuelle ultrasonore dense et la simulation physique. Nous avons aussi proposé un nouveau critère de similarité spécifique à l'imagerie échographique basé sur une étape de détection d'ombre. Enfin, la dernière contribution est liée à une stratégie de suivi innovante permettant d'aligner une image courante avec une image de référence dynamique.

Malgré les résultats prometteurs liés à ces contributions, d'autres investigations permettraient de les enrichir et d'en faciliter l'usage en routine clinique. En effet, l'approche repose actuellement sur une segmentation manuelle. Ainsi, une telle méthode ne peut être utilisée en pratique car cette étape requiert un temps d'interaction important. D'autres approches de segmentation peuvent être utilisées telles que des méthodes semi-automatiques et automatiques (Heimann et al., 2009). Alors que la méthode de suivi a montré d'excellents résultats, une perspective intéressante serait d'évaluer sa performance durant l'intervention. En effet, la nécrose des tissus via l'aiguille peut influencer la précision du suivi car elle génère des variations d'intensités liées à l'apparition de zones hyper-echogeniques et de zones d'ombres. Afin de pallier ce problème, Seo et al. (2011) a proposé une approche permettant d'utiliser ces zones comme des points de repères au cours du suivi. Dans cette thèse, nous avons montré que l'approche montre certaines limites lorsque les structures anatomiques subissent de fortes déformations. Afin de contourner cette limitation, une première solution peut consister à remplacer le modèle masse-ressort-amortisseur par un modèle plus réaliste (Haouchine et al., 2013). Une autre solution serait d'adapter dynamiquement les paramètres mécaniques de ce modèle grâce aux données élastographiques extraites à partir de l'imagerie échographique. Une autre limitation est liée au temps de calcul qui ne permet pas de fournir une méthode temps-réel sur des systèmes échographiques rapides. Pour pallier ce problème, une approche intéressante pourrait considérer des méthodes d'optimisation plus performantes. Par exemple, Delabarre and Marchand (2014) ont proposé une méthode de suivi déformable performante où les déplacements sont calculés dynamiquement grâce à un algorithme de composition inverse permettant un fort gain de temps de calcul.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

This thesis, entitled "Real-time Tracking of Deformable Targets in 3D Ultrasound Sequences" has been performed in the Augmented Healthcare Laboratory of IRT b<>com. It has also involved the collaboration of several partners including the University Hospital Centre of Rennes and IRISA/Inria research institutes. In this work, we propose several contributions that allow estimating the position of anatomical structures in ultrasound imaging over the time. These contributions rely on the combination of methods from physical simulation, computer vision, and medical image processing.

Liver cancer is the sixth most common type of cancer with approximately 800 000 deaths annually worldwide (Torre et al., 2015). It is generally caused by hepatitis B, hepatitis C, or excessive alcohol consumption. While surgery is still the recommended treatment, most of the patients are not surgical candidates. To cope with that issue, minimally-invasive therapies are promising as they aim at locally eliminating malignant lesions from needle insertion. Examples of such interventions are Radiofrequency Ablation (RFA), Cryotherapy and Percutaneous Ethanol Injection (PEI). These therapies have the advantage of reducing patient pain, risk of infection, and length of hospital stay. They can be guided by different imaging modalities such as Ultrasound (US), Computed Tomography (CT), Magnetic Resonance (MR), or Fluoroscopy imaging. However, US imaging is the most used since it provides both low-cost and real-time visual feedback. Despite the benefits of minimally-invasive therapies, their efficiency still depends on the accurate needle placement with respect to targeted structures (Chagnon et al., 2001).

To improve the precision of these therapies, several research groups focused on computer-assisted interventions that involve the processing of multi-modal patient data. This area has gained significant importance over recent years as it allows obtaining more accurate

diagnosis and delivering treatment with greater surgical precision. Therefore, a wide variety of interventions has been impacted by this research field. A few examples of major contributions closely related to our work are highlighted below:

- **Multi-modal Image Fusion:** To improve surgeon visualization, multi-modal registration approaches have been proposed in order to fuse images acquired from different imaging devices. Such methods allow obtaining high resolution and high contrast images. For example, Kadoury et al. (2012) proposed an approach that registers ultrasound images with magnetic resonance data in order to ensure the correct placement of the needle in the liver. Wein et al. (2008b) introduced a method that allows automatic fusion between CT and US imaging in order to improve percutaneous needle biopsy or ablation.

- **Motion Estimation:** To facilitate diagnosis and placement of medical tools, tracking methods have been proposed since they allow following the positions of different structures in consecutive medical images. For example, Baumann et al. (2012) introduced a clinical protocol for prostate biopsy assistance based on the fast and accurate estimation of prostate tissue motion in 3D ultrasound images. Salles et al. (2012) proposed to quantify heart deformation from phase based optical-flow method in order to detect cardiovascular diseases.

- **Robotized Interventions:** Several methods have been proposed in order to improve treatment by automatically guiding medical tools with robotic systems. For instance, Seo et al. (2011) used a motion compensation method for high-intensity focused ultrasound (HIFU) that allows keeping the target and the treatment beam aligned. Wei et al. (2005) proposed a 3D transrectal ultrasound guided robotic assisted system for prostate brachytherapy.

The objective of this thesis consists in improving the surgeon guidance during minimally-invasive therapies for liver tumor ablation. As these interventions are generally ultrasound-guided, we proposed several contributions that allow tracking the motions of anatomical structures in 3D US images. This chapter is organized as follows. In Section 1.1, we describe the challenges related to ultrasound-guided minimally-invasive therapies for liver tumor ablation. Then, Section 1.2 aims at providing a brief description of each key contribution of this thesis. Finally, the thesis organization is detailed in Section 1.3.

1.1 Challenges

In minimally-invasive therapies for tumor ablation, the main objective is to position medical tools with respect to targeted clinical structures in order to produce tissue necrosis. These procedures are generally guided by intra-operative ultrasound imaging in order to detect the positions of different lesions in real-time. However, several problems may perturb the treatment success since they induce incorrect positioning of medical instruments regarding targeted structures:

- **US imaging shortcomings:** Ultrasound imaging has the advantage to be low-cost, portable and real-time contrary to Computed Tomography (CT) or Magnetic Resonance (MR) imaging. However, such imaging system proposes limited field of view due to small acoustic windows, and therefore reduces considerably the surgeon visualization. Additionally, targeted structures and medical instruments are not always clearly visible in US images due to different sources of artifacts, e.g speckle noise, shadows and mirroring effects. Such issues may cause the destruction of healthy tissues due to mis-estimation of targeted structures localization.

- **Target motions:** During the intervention, the targeted structures may undergo large motion that complicates the needle placement. These motions can be caused by physiological motions, because liver is an organ affected by respiratory motion. In addition, medical tools manipulations, such as ultrasound probe pressure and needle insertion, may also generate the deformation of soft tissues. Such problem might force the surgeon to readjust the needle position over the time, and has therefore a significant impact on both the intervention time and on the positioning accuracy.

- **Required accuracy**: Contrary to surgery, minimally-invasive therapies for tumor ablation do not allow treating large regions. Therefore, the intervention has to be accurate in order to be effective. For example, in RFA applications, a particular attention is given to the placement of the needle tip because multiple treatment sessions are required if the target size exceeds 3 cm (Chagnon et al., 2001). Such accuracy constraint may have strong impact on the therapy success.

To solve the different issues, we propose several contributions that can open novel perspectives in computer-assisted interventions based on US imaging and where deformable organs and structures are involved. These contributions aim at improving the surgeon visualization during ultrasound-guided minimally invasive therapies.

1.2 Contributions

The contributions of this thesis are organized into two parts. In the first part of this work, we propose a US tracking method on 3D ultrasound sequences. The contributions on this part include:

- **A piece-wise affine mechanical-based approach:** The proposed method aims at accurately tracking the displacements of a deformable target in ultrasound images. This method consists in three steps. The first step generates a mesh model associated to the target. Once this model is defined, we estimate the target motions over consecutive images or volumes. For this purpose, the model displacements are computed by iteratively summing the internal displacements estimated from a mechanical component, and the external displacements computed from an intensity-based approach. We also propose an adaptation that considers only the information acquired on the scanlines of the ultrasound probe. This approach allows avoiding too much interpolation by considering the ultrasound raw data.

- **An in-depth validation and robustness evaluation:** The method is first validated on simulated ultrasound data by using simulated rigid and non-rigid motions. Then, our method is evaluated on phantom data where the ground truth is provided by odometry and manual annotations. We also performed a comparison of our method with respect to state-of-the-art techniques on real-data by using the 3D databases provided by MICCAI CLUST'14 and CLUST'15 challenges. Thus, the proposed approach has been tested by tracking 34 different anatomical features from 22 3D US sequences. The ground truth is given by using the manual annotations of landmarks from three experts of target positions over each frame. Finally, several experiments are performed in order to show the method robustness regarding mechanical model parameters, segmentation error, and model mesh quality.

The second part of this work focuses on new image similarity criterion and tracking strategy that we proposed. The contributions on this part include:

- **An ultrasound specific similarity criterion:** The proposed criterion aims at correctly tracking structures that are affected by different ultrasound shortcomings including imaging gain change, shadows and beam-angle variations. This criterion improves an existing criterion called "Sum of Conditional Variance" (Pickering et al., 2009) by minimizing the influence of shadowed image regions. The method consists in combining the computation of the joint probability density function of the image intensity with the quality measurement of US images. The proposed criterion has the advantage to

be invariant to local and global intensity shifts and is not computational demanding. Experimental results reveal that the new criterion shows good performance even if large shadows significantly occlude the target or large ultrasound gain modifications are applied.

- **A hybrid tracking strategy:** Tracking methods can be divided into pairwise and groupwise tracking methods. The first type of approach consists in comparing the current image with a selected image, while the second registers the current image with all previous images. The main drawback of pairwise tracking methods is that their performance strongly depends on the quality of the selected image. The second approach solves this issue at the expense of a high computational cost for long US sequences. We propose an hybrid method that consists in improving the selected image if its quality is not good enough. To do so, unconfident voxels of the selected image are replaced by voxels of the consecutive images if their associated confidence is greater. Experimental results show that our method is more accurate than pairwise strategies with the same computation time.

1.3 Thesis organization

Chapter 2 aims at presenting the medical context associated to the thesis. Section 2.1 gives background information about the liver including its anatomy, its physiology, and its pathology. Section 2.2 then describes the medical imaging technologies that allow representing the liver for clinical analysis and medical interventions. Section 2.3 describes the minimally invasive therapies for treating liver diseases.

Chapter 3 provides background information in order to understand this thesis and its contributions. Section 3.1 describes the main components for both tracking methods and registration algorithms. Section 3.2 focuses on the existing approaches in the literature for target tracking in ultrasound sequences, and for registering ultrasound imaging of the liver to pre-operative CT/MRI imaging.

Chapter 4 describes the first part of our contributions, dedicated to the ultrasound tracking method based on the combination of a mechanical approach and an intensity-based approach. Section 4.1 and 4.2 describe the proposed method and its performance for 3D ultrasound sequences. Finally, Section 4.3 demonstrates the method robustness regarding target elasticity, target velocity and mesh properties.

Chapter 5 focuses on the second part of our contributions that includes a novel ultrasound-specific criterion as well as an hybrid tracking strategy. Section 5.1 describes the proposed Sum of Confident Conditional Variance (SCCV) criterion and compares its performance

regarding both classical and confidence-based similarity measures. Section 5.2 provides a description of the proposed hybrid tracking strategy combining the existing pairwise strategy with quality measurement of US images. To demonstrate the performance of the novel strategy, the proposed method is tested on simulated data and phantom data.

Chapter 6 describes our last work, which focuses on multi-modal application in order to improve the surgeon visualization. In this chapter, we propose a specific protocol that allows registering 3D US imaging and IRM acquired from University Hospital Center of Rennes.

Chapter 7 concludes this document. The current issues and thesis objectives are first reminded in Section 7.1. In Section 7.2, we summarize how the proposed contributions help to solve the issues. The perspectives and potential improvements are finally addressed in Section 7.3.

Finally, appendix A lists the publications and patent applications published during the thesis.

Chapter 2

Medical context

This chapter aims at describing the medical context of the thesis in order to better understand the motivations and the issues associated to percutaneous minimally-invasive therapies of the liver. The liver is responsible for many vital functions and liver cancer has a significant impact on the survival rate of the patient. Depending on the stage and the speed of tumor growth, several treatment options may be proposed including surgery, tumor ablation, radiation therapy and chemotherapy. Although surgery is still the most efficient treatment, the majority of the patients are not surgical candidates since they may have not enough healthy liver to cope with the resection. Indeed, there is a risk that the cancer could come back in the future after a liver resection. this thesis, we focus on minimally-invasive therapies for tumor ablation as they represent safe and effective treatment alternatives. As we focus on minimally-invasive therapies for tumor ablation, the other treatment options of the liver cancer, e.g. resection, transplantation, and chemotherapy, are not reviewed in this chapter. This latter is organized as follows. Section 2.1 focuses on the liver organ by describing its macroscopic anatomy, microscopic anatomy and its pathology. Section 2.2 describes the different imaging modalities that may be used in order to guide interventions or to determine the diagnosis. As each imaging system has strength and weakness, we also discuss about the performance of each system. Section 2.3 aims at providing a description of the different minimally-invasive interventions that allow treating liver tumors. We also discuss about current limitations associated to these treatments.

2.1 Liver

In biology, an organ is a collection of tissue joined in a structural unit that allows serving a common function. The main function of the liver consists in secreting bile that aids the digestion of lipids in the small intestine. However, many others vital functions are performed

by the liver such as the filtration, synthesis of bio-medical substances, and storage of chemical compounds. The liver tissue contains mainly hepatic cells, also called hepatocytes, which are responsible for liver functions. However, it is also composed of others cells, including sinusoidal hepatic endothelial cells, Kupffer cells and hepatic stellate cells, that have the ability to regenerate the liver.

2.1.1 Macroscopic anatomy

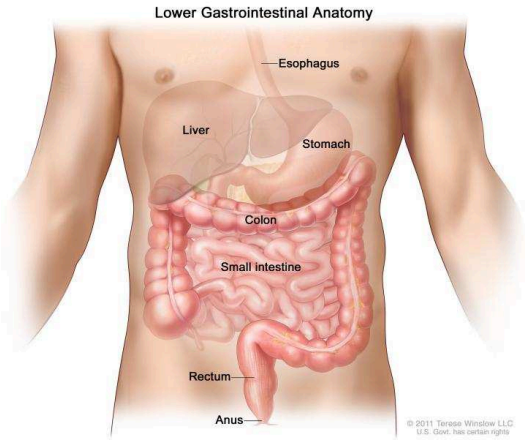


Fig. 2.1 Lower gastrointestinal anatomy.[1]

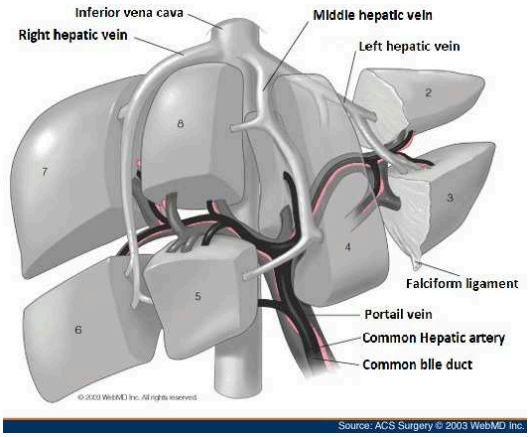


Fig. 2.2 Couinaud Classification.[2]

The liver is the largest organ of the human digestive system with brown color and smooth surface. It has three surfaces including superior, inferior and posterior surfaces. The superior surface is linked to the diaphragm by the falciform ligament. The inferior surface is in relation with the stomach and the duodenum, while the posterior surface is in direct contact with the diaphragm. As illustrated in Fig. 2.1, the liver is located in the upper-right portion of the abdominal cavity and is composed of two main lobes. From the Couinaud classification shown in Fig. 2.2, we can observe that the liver anatomy can be divided into eight functionally independent segments delimited by hepatic venous. As each segment has its own vascular flow, an entire segment can be removed without damaging the vascular system during resection intervention. The liver is responsible for over 20 percent of the oxygen consumption, and is considered as a portal organ because it is located between the portal circulation and the vena cava.

[1] *http://www.anatomy-diagram.info/normal-anatomy-of-the-liver/*
[2] *http://www.intechopen.com/books/hepatic-surgery/liver-resection-for-hepatocellular-carcinoma*

2.1.2 Microscopic anatomy

At microscopic scale, the liver is organized into lobules composed of millions of hepatics cells linked together. Each lobule has hexagonal shape centered on a central vein corresponding to a branch of hepatic vein. Each lobule corner allows supplying blood to the lobule from branches of the hepatic artery and the portal vein. The blood is then mixed in the hepatic vessels and routed out to the central vein branch. Inversely, the bile is synthesized and is drained away from the central vein toward bile duct branch. Then the bile is routed out of the liver via the common hepatic duct leading to the intestine. Due to the number of liver functions, this organ is involved in a variety of pathological processes that can affect the microscopic appearance of the liver. In this thesis, we focus only on primary and secondary liver cancer.

2.1.3 Liver cancer

Liver cancer, also known as hepatic cancer, corresponds to single or multiple tumors that spread into the liver affecting the hepatic cells with different growth patterns. Two main categories of cancer may affect the liver including primary and secondary liver cancers.

Primary cancer

Primary liver cancer corresponds to cancer that starts from the liver. It is mainly caused by cirrhosis that may be induced by several factors including:

- **Alcohol consumption:** During excessive alcohol consumption over a long period of time, normal liver functions may be interrupted. Such problem leads to the destruction of liver cells. It is worth mentioning that certain individuals are more affected by alcoholic cirrhosis, because alcohol metabolization of the liver depends on several parameters, including gender, age, and weight.

- **Viral hepatitis:** Several forms of hepatitis viruses may attack the liver and may cause liver inflammation and cirrhosis. The two most commons form of viral hepatitis are hepatitis B and C that are contracted by infected blood and body fluids.

- **Genetic disorders:** Inherited disorders, such as Haemochromatosis and Wilson's diseases, may cause cirrhosis as they induce accumulation of toxic substances in the liver. For example, Haemochromatosis is a inherited disease causing an excessive amount of iron in the body. With Wilson disease, the mutation of an abnormal gene introduces accumulation of copper in the tissues.

- **Autoimmune disorders**: Immune system disorders may also cause cirrhosis. Examples of autoimmune liver diseases are primary biliary cholangitis (PBC), and autoimmune hepatitis. PBC is responsible for destruction of the small bile ducts within the liver. Such disease therefore limits the bile flow inducing progressive inflammation, scarring, and the toxic effects of accumulating waste products. Autoimmune hepatitis is another immune disorder, where the immune system considers hepatic cells as foreign and destroys them.
- **Non-alcohol related fatty liver disease (NAFLD)**: NAFLD occurs when fat is deposited in the liver tissues due to other causes than excessive alcohol consumption. Several factors may introduce such effect including medication consumption and soft drinks containing high concentration of fructose.

Secondary cancer

The secondary liver cancers, also called cancer metastases, spread into the liver from another part of the body. This is caused when cancer cells break away from the primary cancer and move to other body parts through the bloodstream or lymph system. It is worth mentioning that secondary liver cancers are more prevalent than primary liver cancers. Secondary liver metastases are generally caused by cancers of the colon, pancreas, stomach, lung or breast.

Liver cancer diagnosis

To determine the presence of liver cancer, several tests may be conducted such as imaging and blood tests. Imaging tests consist in finding suspicious areas by using Ultrasound, Computed Tomography or Magnetic Resonance imaging. The different technologies of these imaging systems are described in the next section. Different bloods tests may also be used since liver cancer modifies levels of certain substances in the blood. These tests may include viral hepatitis tests, Alpha-fetoprotein blood tests.

2.2 Medical imaging

2.2.1 Computer tomography

Computer tomography, also called CT, refers to computerized X-ray imaging that produces cross-sectional images of the patient. It is based on the principle that the tissue density, passed by an X-ray beam, can be retrieved from the measurements of X-ray attenuation. Such

imaging system therefore consists in taking several X-rays images obtained from different angles. Image reconstruction algorithms allow obtaining 3D CT volume that shows the skeleton, organs, and soft-tissue of the patient. The first CT device has been created in 1972 by Godfrey Newbold Hounsfield and Allan MacLeod Cormack. The original prototype was able to generate CT image of the brain in less than 2.5 hours. Several improvements have been performed regarding the speed acquirement, image quality, and image resolution. Today, new generations of scanner allow taking CT image in less than 1 second.

Acquisition system: CT scanners are composed of two elements including an X-ray source and X-ray sensors. The source allows generating the X-ray beam that passes through the patient tissue, while the X-ray sensors measure their X-ray attenuation coefficients. X-ray sensors are positioned on the opposite side from the X-ray source, and they rotate around the studied patient in order to measure the attenuation of the whole volume.

Image characteristics: CT image has the advantage to provide images of high quality. Furthermore, contrast materials may be used in order to highlight structures such as blood vessels that would be difficult to delineate from their surroundings tissues. However, CT system exposes the patient to an amount of radiation that may produce adverse effects.

2.2.2 Magnetic resonance imaging

The Magnetic Resonance Imaging, also called MRI, refers to a medical imaging technique that allows imaging the human body. It relies on the physical principle of Nuclear Magnetic Resonance discovered by Isidor Rabi in 1938. This principle induces that hydrogen atoms in a magnetic field may absorb and emit radio-frequency (RF) signals. Magnetic field allows aligning protons within the hydrogen atoms, while radio-frequency waves knock them from their positions. The re-alignment of the protons produces radio-frequency signal that can be measured by MRI acquisition system. As these protons are re-aligned at different rates depending on the tissue density, the emitted RF signal is also different. This characteristic therefore allows distinguishing and delineating the different tissues in order to create an image. The first MRI medical device was created in 1969 by Paul C. Lauterbur.

Acquisition system: The MRI acquisition system is composed of three elements including magnet, coils, and R.F antenna. The magnet allows generating a constant and permanent magnetic field across the studied volume. The coils introduce linear variation of the magnetic field along specific axis and therefore allow retrieving the spatial direction of the RF signal.

Finally, the antennas allow generating the radio-frequency signal that is absorbed and re-emitted by the hydrogen atoms.

Image characteristics: The MRI imaging has the advantage to present good imaging quality that depends on the magnetic field strength. The greater is the magnetic field intensity, the better is the imaging quality. Contrast agents may be injected in order to improve the contrast between different tissues. Furthermore, the contrast of different anatomical structures or pathologies can be emphasized by exploiting the different relaxation properties of the hydrogen atoms. By doing so, different MRI imaging types can be obtained such as T1-weighted imaging and T2-weighted imaging.

2.2.3 Ultrasound imaging

Ultrasound imaging, also called ultrasonography, is a intra-operative imaging technique that produces either 2D or 3D images of the human body. It relies on the physical principle that ultrasound waves are partly reflected whenever a change in acoustic impedance is encountered between two media. Therefore, we can determine the limits between soft-tissues by emitting ultrasound waves through tissues and measuring their reflection. US waves correspond to acoustic waves whose frequency is higher than those audible by human. They have been discovered by Spallanzi in 1794 by deducing that bats use ultrasound in order to navigate by echolocation. In 1880, ultrasonic waves have been reproduced by Pierre Curie when he discovered the piezo-electric effect. However, its first application for medical purpose was proposed by Dussik in 1941 in order to obtain brain imaging.
Acquisition system: The ultrasound probe, also called transducer, is the main element of an ultrasound system and has two functions. Its first function consists in emitting US signal through human tissues, while the other function allows receiving the US signal reflected by tissues interfaces and converts it into electric signal. This transformation is performed from piezo-electric elements attached to the ultrasound probe. The image is formed by analyzing both the time and the strength ratio between the emitted and the reflected ultrasound wave.

Image characteristics: The ultrasound systems have the advantage to render real-time images with non-ionizing energy. Furthermore, they are relatively affordable compared to other imaging systems. Consequently, this imaging modality is very well suited for interventions requiring image guidance. However, as illustrated in Fig. 2.3, US imaging may produce bad quality images due to the presence of bones or gas as they have high acoustic

[3] *http://www.prweb.com/releases/2011/5/prweb8377014.htm*

(a) (b)

Fig. 2.3 (a) Illustration of US machine and probes[3]. (b) Illustration of 2D US image.

impedance. In addition, speckle noise can also perturb significantly the visualization due to the interference of the returning wave at the transducer aperture.

2.3 Percutaneous minimally-invasive therapies

In this section, we describe the main minimally-invasive therapies used for hepatic tumor treatment including radiofrequency ablation, cryotherapy, and percutaneous alcohol ablation. These interventions are generally guided by ultrasound imaging since they require accurate placement of needle with respect to anatomical structures. It is worth mentioning that others invasive treatments can be used instead. For example, resection is considered as the most curative treatment but is feasible for only 20 % of the patients. Liver transplant can also be recommended for end-stage liver disease. However, this intervention is not always carried out due to the small number of available liver donations. Chemotherapy is used to disrupt the cancer growth, or reduce symptom, but can not completely cure liver cancer.

2.3.1 Radiofrequency ablation

The radiofrequency ablation (RFA) is a treatment technique that is used to destroy tumors in the liver. During this intervention, a needle is inserted within the tumor, and delivers a high frequency alternating electrical current that generates heat burning out the adjacent tissue. As illustrated in Fig. 2.4, this treatment causes necrosis of the tissue whose the diameter around

the needle increases with intervention time. Generally, the surgeon has to maintain in place the needle for about 10 to 15 minutes per nodule.

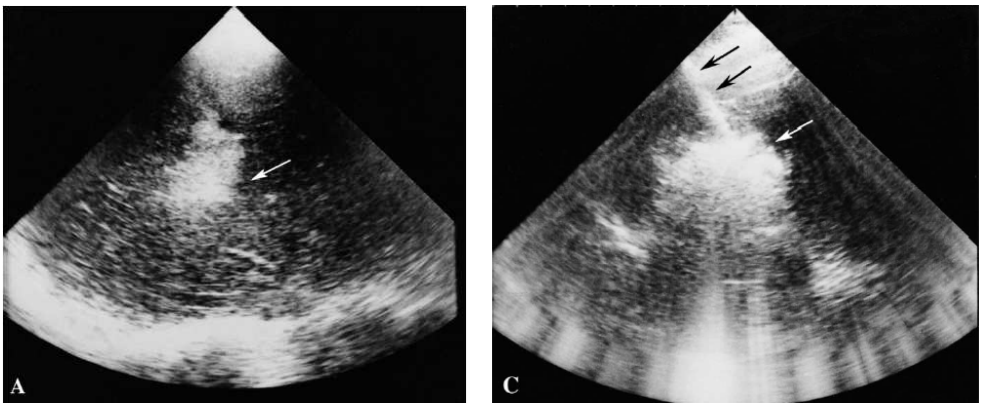


Fig. 2.4 Illustration of radio-frequency ablation procedure guided by ultrasound imaging (Chagnon et al., 2001). (Left) image representing liver tumor before treatment. The hyperechoic area represents liver metastases. (Right) image representing liver tumor during treatment. The RFA needle is inserted (black arrows). The hyperechoic is larger due to the necrosis of tissues.

2.3.2 Cryotherapy

The cryotherapy is a local percutaneous treatment used to kill malignant cells. The principle of this technique consists in destroying the tumor by using a cryoprobe that delivers low temperatures. Argon gas, or nitrogen liquid are the most commonly used agents to destroy the tissue. The intervention time is estimated to 20 minutes per nodule. Regarding Chagnon et al. (2001), the complications are more important with the cryotherapy compared to the RFA described in the previous section.

2.3.3 Percutaneous ethanol injection

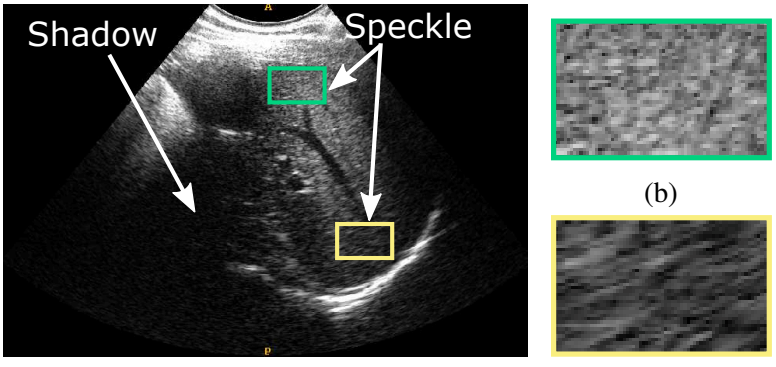
The Percutaneous Ethanol Injection (PEI) is a local percutaneous treatment that injects ethanol directly into the malignant tissue of the liver. The alcohol kills the malignant cells by dehydrating the tissue, and stopping its blood supply. This method is well suited when the patient has a small number of little tumors. Regarding Chagnon et al. (2001), the RFA intervention shows better results than the PEI.

2.3.4 Current limitations

Minimally-invasive therapies for tumor ablation are increasingly used since they reduce risks of complication, patients pain, and length of hospital stays. Despite the benefits of these treatments, there are still some unresolved technical limitations that reduce their accuracy and their effectiveness. In this section, we detail different types of limitations related to imaging shortcomings, treatment characteristics, and soft-tissue movements.

Ultrasound imaging shortcomings

As stated before, the minimally-invasive interventions are generally performed under ultrasound guidance. However, several US imaging shortcomings may affect the visualization of the surgeon. Among these shortcomings, ultrasound shadow is one of the most perturbing since it may occlude targeted structures during the procedure. This artifact is caused by strong reflections of ultrasound waves due to bad contact between skin and probe, or gas production during the ablative process (Künzli et al., 2011). Furthermore, US shadows can also be introduced by solid structures such as bones or calculus. Another common drawback of US imaging is related to speckle noise introduced by the presence of micro-structures within soft-tissues. This noise generates granular intensity variation in the sequence of ultrasound images and complicates the accurate localization of medical tools such as the needle tip position. Finally, it is worth mentioning that ultrasound imaging has small field of view and limited acoustic window due to chest size and rib spaces of patients. As a consequence, the structures of interest may be only partially visible. In Fig. 2.5, we illustrate different US shortcomings.



(a) (b) (c)

Fig. 2.5 Illustration of ultrasound shortcomings. (a) US image that represents partial view of the liver. (b-c) Speckle noise illustrations.

Treatment characteristics

The success of minimally-invasive therapies can also be affected by issues related to treatment characteristics. Indeed, these interventions can only treat small regions around the needle tip. Higgins and Berger (2006) provide a study that shows that the tumor can be completely removed if its size is less than 5 cm. As a consequence, the size of malignant tissues may have an impact on the treatment effectiveness. The tumor localization may also play a significant role since minimally-invasive thermal procedures, such as RFA or cryotherapy, can not be applied to liver regions close to large vessels. These regions are protected by heating and cooling effects provided by the temperature of the blooding flow. To cope with that issue, percutaneous ethanol injection can be used instead but its effectiveness depends on the heterogeneity of alcohol distribution within the soft tissues.

Dynamic structures

During minimally-invasive therapies, the accurate needle placement can be perturbed by the dynamics of soft-tissues that cause displacement of targeted structures over the time. These motions are generally caused by physiological movements such as respiratory or cardiac motions. However, medical tools manipulation, such as ultrasound probe pressure on skin, may also introduce displacements and deformation of soft tissues. Such issue makes the surgeon readjust the needle position in order to ensure success of the therapy and avoid eliminating healthy tissues. The problem becomes even worse if the acquisition rate of the US probe is not sufficient in order to ensure a continuous visualization of soft tissue displacement.

2.4 Conclusion

In this chapter, we described the medical context of the thesis related to mini-invasive interventions for liver cancer treatment. Compared to classical surgery, these therapies are promising since they significantly reduce pain and recovery time of the patient. However, the effectiveness of these procedures depends on their accuracy since these interventions rely on highly localized surgery. To improve their precision, several research groups proposed different approaches that allow determining the position of targeted structures over the time. In the following chapters, we describe the different components and the principle of these approaches. We also propose several novel contributions that aim at improving both accuracy and robustness of the tracking.

Chapter 3

Related work

During minimally-invasive therapies of the liver, such as RFA, an accurate positioning of needle is required in order to treat malignant tissues. However, such task may necessitate considerable time and efforts of the surgeon due to motions of structures of interest. Furthermore, the visualization of different tissues may also be perturbed by different ultrasound imaging artifacts described in the previous chapter. To cope with that issue, registration and tracking methods may be employed as they allow determining both the displacement and the deformation of targeted structures between two images.

This chapter describes the different components and related work associated to registration and tracking methods. Only approaches dedicated to US images are reviewed since this imaging modality is the most used technology in order to guide minimally-invasive percutaneous interventions. We also detail the performance of different registration methods that allow aligning US imaging to other pre-operative imaging.

The chapter is structured as follows. In Section 3.1, we give an overview of the main components of registration techniques. In Section 3.2, we detail some work related to the tracking and registration techniques. Section 3.3 concludes the chapter by detailing possible future research directions.

3.1 Tracking/Registration components

The objective of registration methods is to define the geometric transformation that aligns reference and moving images. The optimal geometric transformation is found by minimizing

the following cost function expressed as:

$$\arg\min_{T} S(I_{t_0}, T \circ I_t) + R(T) \tag{3.1}$$

where I_{t_0} and I_t represent respectively the intensity vector of the reference image and the moving image. Throughout this thesis, we interchangeably use the terms moving image and current image. $T \circ I_t$ denotes the intensity vector of the aligned image representing the moving image transformed by the geometric transformation T. Any registration technique also depends on:

- **Similarity criterion term S:** The similarity criterion allows measuring the alignment between the reference image and the current image. Ideally, this criterion has a single global optimum when both images are aligned.
- **Transformation Model T:** The transformation model aims at representing the space of possible displacement ensuring alignment between two images. Transformations are generally categorized into rigid and non-rigid models.
- **Regularization term R:** The objective of the regularization term is to penalize unrealistic properties of the transformation, such as non smooth deformations, target volumes changes and folding.
- **Optimization strategy:** The optimization strategy allows finding the optimal parameters of the transformation that aligns the reference image to the current image by minimizing the cost function described above.

The tracking approaches also depend on these components because they are strongly related to registration methods. Indeed, the tracking objective is to define the geometric transformation relating the target position over a set of consecutive images. The major difference is that tracking methods generally require fast algorithms in order to handle real-time image stream. It is worth mentioning that tracking and registration techniques may also be categorized regarding several other criteria, e.g. imaging modality, type of anatomical structure and type of user interaction. However, they are not discussed in this section since it is out of the scope of this thesis.

3.1.1 Matching criteria

The matching criterion quantifies the level of alignment between a reference image and a moving image. They are often categorized into intensity-based and feature-based matching

criteria (De Luca et al., 2015). The first class aims at determining the alignment between images from intensity correspondence measurements. These methods have the advantage to provide good accuracy and robustness. However, they are generally computationally demanding due to the high number of pixels or voxels in the image. Feature-based methods are based on the extraction of features, e.g. points, edges, local intensity statistics, between reference and moving images. Therefore, their matching criteria are generally formulated as distance between the feature locations. Feature-based methods are generally faster as they do not use the intensity of the whole image. However, the main drawback is they may provide incorrect results when features are not well detected.

Sum of Squared Distance (SSD)

In tracking and registration applications, the Sum of Squared Distance criterion is the most used criterion due to its simplicity (Yeung et al., 1998). It can be expressed as follows:

$$S_{SSD} = (I_{t_0}(\mathbf{p}_{t_0}) - I_t(\mathbf{p}_t))^2 \tag{3.2}$$

where I_{t_0} and I_t are vectors that represent respectively the image intensity at time index t_0 and at time index t. $\mathbf{p}$ and $\mathbf{p}_{t_0}$ represent respectively the considered points positions in the images I_{t_0} and I_t. As it can be seen from previous equation, the SSD cost function directly represents the squared difference between reference intensity and current intensity. Consequently, this matching criterion is well suited when the intensity of physical point is constant between the reference and current images.

Sum of Absolute Distance (SAD)

As proposed by Touil et al. (2010), some work are also based on the use of Sum of Absolute Distance in order to define the intensity correspondence between images. This criterion can be expressed as follows:

$$S_{SAD} = |I_{t_0}(\mathbf{p}_{t_0}) - I_t(\mathbf{p}_t)| \tag{3.3}$$

As it can be seen from previous equation, the SAD cost function directly represents the absolute difference between two images. As for SSD, this criterion is adapted when the intensity of interest structure remains constant over the time. SAD cost function provides better results than SSD in the presence of Laplacian noise. However, SSD provides more accurate results than SAD with Gaussian noise.

Normalized-Cross Correlation (NCC)

As it can be seen from previous equations, SSD and SAD criteria are well-suited if there is no intensity shift between reference and current images. However, this assumption is not always valid. To cope with this issue, cross-correlation criteria, assuming linear relationship between the intensity distributions, have also been proposed (Irani et al., 1992). A typical example is the normalized cross-correlation criterion that can be expressed as follows:

$$S_{NCC} = \frac{(I_{t_0}(\mathbf{p}_{t_0}) - \widehat{I_{t_0}})(I_t(\mathbf{p}_t) - \widehat{I_t})}{\sigma_{I_t}\sigma_{I_{t_0}}} \tag{3.4}$$

where σ_{I_t} and $\sigma_{I_{t_0}}$ represent respectively the standard deviation of the images I_t and I_{t_0}. $\widehat{I_t}$ and $\widehat{I_r}$ represent the average intensity of the images I_t and I_{t_0}.

Mutual Information (MI)

As the intensity relationship is not necessarily linear, more sophisticated approaches based on theoretical measurements from probability theory and statistics have been studied. These criteria are generally well-suited for multi-modal application, where images are acquired from different imaging devices with specific physical properties. Among them, the mutual information, as proposed by Elen et al. (2008), has been the most widely used similarity criterion. It measures the amount of information that can be obtained about one variable by observing another. This gives an indication about the quantity of information that is shared between two variables. The mutual information criterion (MI) can be expressed as follows:

$$S_{MI} = \sum_{k=1}^{N_v} p_{I_t I_{t_0}} \log\left(\frac{p_{I_t I_{t_0}}}{p_{I_{t_0}} p_{I_t}}\right) \tag{3.5}$$

where $p_{I_{t_0}}$ and p_{I_t} are respectively the probabilities associated to the reference and current images. $p_{I_t I_{t_0}}$ denotes the joint probability density function between reference and current images normalized by the number of pixels.

Feature-based distance

In feature-based registration approaches, the matching is performed by measuring the distance between feature locations. It is worth mentioning that feature-based methods may also use matching criterion in order to determine the correspondence between features from reference image to moving image. This is performed by comparing the pairwise distances between

feature descriptors that allow characterizing feature. To do so, several distance measurements can be used such as:

- **Euclidean distance:** It corresponds to the norm of the difference of vectors expressed as follows:

$$D_e = ||\mathbf{x} - \mathbf{y}||^2 \tag{3.6}$$

  where $\mathbf{x}$ and $\mathbf{y}$ represent vectors of the feature locations in the reference image and in the current image. As it can be seen from previous equation, all the vector components contribute equally to the Euclidean distance.

- **Mahalanobis distance:** it corresponds to another distance that takes into account the variability between two vectors (Mahalanobis, 1936). It can be expressed as follows:

$$D_m = ||(\mathbf{x} - \mathbf{y})^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})||^2 \tag{3.7}$$

  where $\mathbf{S}$ represents the covariance matrix between descriptor vectors of the features in the reference image and in the current image.

3.1.2 Transformation model

The transformation model allows defining spatial correspondence by mapping point locations from reference image to moving image. In the literature, different transformation models have been proposed in order to represent several type of displacements from rigid to non-rigid motions. Therefore, different numbers of parameters may help to characterize transformation models.

Rigid/Affine transformation

A number of tracking approaches, as proposed by (Masum et al., 2014), are based on rigid model since it provides constant displacement field without any deformation. It can be expressed by translation and rotation parameters as follows:

$$T(\mathbf{p}_{t_0}) = \mathbf{R}\mathbf{p}_{t_0} + \mathbf{t} \tag{3.8}$$

where $\mathbf{R}$ and $\mathbf{t}$ represent respectively the rotation matrix and translation vectors. $T(\mathbf{p}_{t_0})$ represents the transformed pixel/voxel positions. Such model is well-suited for rigid anatomical objects such as bones. Affine transformation, as proposed by Wein et al. (2008a), slightly differs from rigid model by replacing the rotation matrix by affine matrix combining shear, scale and rotation parameters.

Block-matching transformation

As mentioned earlier, the rigid transformation model can not correctly represent deformable structures. To cope with that issue, block matching transformation models, introduced by (Jain and Jain, 1981), can be used since they allow estimating the displacements by subdividing the whole image into small image blocks. Then, the displacement field is obtained by computing the translation displacement of each block. As illustrated in Fig. 3.2, deformable implementation of block-matching transformation models can also be designed by smoothing the displacement field between the blocks (Basarab et al., 2008).

Thin plate splines

The thin plate splines (TPS), as proposed by Bookstein (1989), is a technique based on the radial basis functions in order to represent non-rigid deformation. This transformation model interpolates the whole displacement field by using a set of control point $\mathbf{x}_l$. It can be expressed as follows:

$$T(\mathbf{p}_{t_0}) = \mathbf{A}\mathbf{p}_{\mathbf{t_0}} + \mathbf{t} + \sum_{i=0}^{N} w_i \phi(||\mathbf{p}_{\mathbf{t_0}} - \mathbf{x}_{l_i}||) \tag{3.9}$$

where ϕ denotes a distance function which is $\phi(r) = r^2 log(r)$ in the 2D case, and $\phi(r) = r$ in the 3D case. The matrix $\mathbf{A}$ denotes respectively the affine transformation combining shear, scale, and rotation parameters. w_i represents the weight of each control point from linear equation system. N represents the number of known control points. The thin plate splines transformation is considered as global transformation since each control has a influence on the whole displacement field of the image.

Free-form deformation

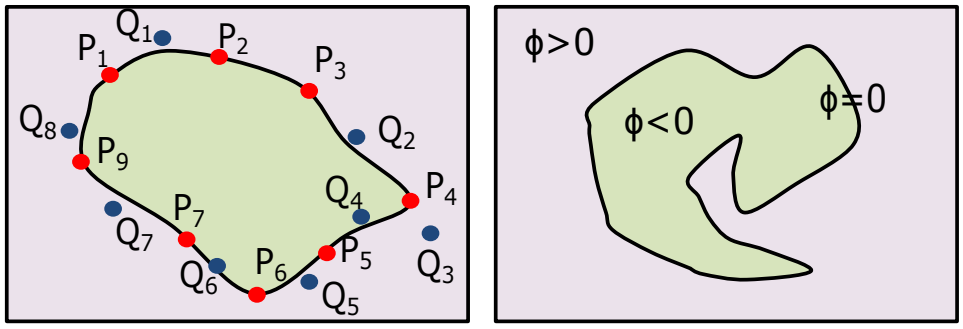
The Free-Form Deformation (FFD), as proposed by Sederberg and Parry (1986), is commonly used as transformation model in medical imaging registration. In contrast to TPS, the displacement of a control point provided by FFD has only local influence on the displacement field, and therefore allows representing highly localized deformation. The free-form deformation defines a rectangular grid of controls points which is superimposed upon the image whose dimension is lower than the image dimension. The displacement field of the whole image expressed in the following equation is then locally interpolated from the known value by using splines functions:

$$\mathbf{u}(\mathbf{x}) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} B_l(u)B_m(v)B_n(w)\mathbf{d}_{lmn} \tag{3.10}$$

where B_l, B_m and B_n denote the B-spline functions. $\mathbf{d}_{lmn}$ is the displacement of a specific node of the grid. The coordinates u, v and w represent the normalized coordinates of specific point within a grid block.

Contour model

The ultrasound tracking methods are also often based on segmentation methods providing transformation model that defines a contour around the region of interest. These transformation models can be categorized into explicit (Nascimento and Marques, 2008) and implicit (Osher and Fedkiw, 2003) warping functions. The explicit representation allows defining the contour from a set of control points and spline equations as shown in Fig. 3.1a. The implicit model represents the contour from signed distance map function around the object. These models have the advantage to be robust to topology changes, intersection and splitting operation. A typical implicit transformation is the level-set framework where the distance map is encoded from a grid of points. As illustrated in Fig. 3.1b, the curve is defined by the isolevel of value 0.



(a) (b)

Fig. 3.1 Explicit and implicit contour models. (a) Spline contour defined by the control points Q_i and passing through the points P_i. (b)Implicit contour model defined by isolevel of value 0. ϕ denotes the signed distance map function.

3.1.3 Regularization

In tracking applications, non-rigid transformation may provide non-physically plausible deformations such as folding, large local stretching or shrinking. To cope with that issue,

regularization terms can be added as post-processing steps. They allow providing realistic tracking results by constraining the deformation field. The most used regularization terms are the following:

- **Volume preservation**: In registration methods, volume preservation regularization term can be used in order to ensure that some visualized tissues are incompressible (Tanner et al., 2000). A simple example is a rigid part of the body such as a bone structure.

$$R(T) = \frac{1}{N_v} \sum_{k=1}^{N_v} |\log(\det(\mathbf{J}_T(\mathbf{p}_t)))| \tag{3.11}$$

  where $\mathbf{J}_T$ denotes the Jacobian transformation representing the spatial derivatives of the transformation. N_v represents the number of voxels affected by the transformation T.

- **Spatial smoothness**: The spatial smoothness constraint, as introduced by Heyde et al. (2012), is based on the assumption that the displacement field between two neighboring voxels does not involve sudden local changes. It can be expressed as follows:

$$R(T) = \frac{1}{N_v} \sum_{k=1}^{N_v} |\nabla u(\mathbf{p}_t)|^2 + |\nabla v(\mathbf{p}_t)|^2 + |\nabla w(\mathbf{p}_t)|^2 \tag{3.12}$$

  where ∇u, ∇v, and ∇w represent the vectors containing spatial derivative of the displacement field.

- **Inverse Consistency**: Interchanging the order between reference and current images may produce slightly different transformations due to the existence of several optimums. Such drawback may significantly impact the algorithm precision and may produce biased results. To cope with that issue, the inverse consistency regularization, as proposed by Christensen and Johnson (2001), allows penalizing inconsistent inverse transformations. This can be expressed as follows:

$$R(T, T_b) = \frac{1}{N_v} \sum_{k=1}^{N_v} ||T_b(T(\mathbf{p}_t)) - \mathbf{p}_t||^2 \tag{3.13}$$

  where T represents the normal forward transformation that aligns reference image to current image. T_b denotes the inverse transformation ensuring alignment from current image to reference image.

- **Time Prediction**: As the ultrasound images are acquired consecutively, several tracking methods assume that the evolution of the transformation model can be retrieved from past images. Therefore, such approaches aim at predicting the optimal transformation from dynamic systems. Generally, the prediction is performed by using either Kalman filter (Kalman, 1960) or Particle filter (Gordon et al., 1993). It is worth mentioning that Kalman filter is well-suited when the relationship between motions and measurements is linear. However, particle filter provides better prediction when this relation is non-linear.

3.1.4 Optimization strategies

The optimization problem aims at finding the optimal parameters by minimizing the cost function defined in the Eq. (3.1). They are generally categorized into continuous strategies and discrete strategies. It is worth mentioning that the first class of methods is the most used. Therefore, we do not review discrete methods as they are rarely used for ultrasound tracking application.

Continuous strategies

The main principle of continuous optimization methods is to estimate the parameters from previous estimation. The update rule can be expressed as follows:

$$\mathbf{q}^{k+1} = \mathbf{q}^{k} + \alpha_i \mathbf{g}^{k}(\mathbf{q}^{k}) \tag{3.14}$$

where $\mathbf{q}^k$ and $\mathbf{q}^{k+1}$ represent respectively the parameters values estimated at iterations k and $k+1$. $\mathbf{g}^k$ represents the search direction obtained from the optimization strategy. α_i denotes the optimization gain. In order to compute the search direction $\mathbf{g}^k$, several methods are proposed including:

- **Gradient descent:** This method, as described by Moré and Thuente (1994), allows considering the search direction to be proportional to the first-order derivative function at the current point such that $\mathbf{g}^k = \nabla_{\mathbf{q}}(\mathbf{q}^k)$.

- **Conjugate gradient descent:** This method, as introduced by Fletcher and Reeves (1964), expresses the search direction as a linear combination of previous calculated first-order derivative functions. Therefore, the search direction can be defined with a linear combination which can be expressed as $\mathbf{g}^k = \nabla_{\mathbf{q}}(\mathbf{q}) + \beta \mathbf{g}^{k-1}$, where β is an optimization gain.

- **Powell's method:** Powell's method, as described by pre (2007), is a zero-order optimization method minimizing the cost function thanks to a set of initial search vectors that depend on the number of parameters. The minimization is performed along each search vector iteratively. At each iteration, the displacement vector can be expressed as a linear combination of the search vectors and is added to the end of the search vector list. When the search has been executed for each N vector, the search vectors are shifted $\mathbf{g_t^i} = \mathbf{g_{t-1}^{i+1}}$. And the last search vector can be computed as $\mathbf{g_t^i} = \mathbf{q}^k - \mathbf{q}_0$. Despite its computational inefficiency compared to others methods, Powell minimization has been used by several authors such as Gee (1999).

- **Newton-type methods :** This method is based on the computation of second-order derivatives by computing the inverse of the Hessian matrix $\mathbf{H}$ in order to determine the update step $\mathbf{g_t}$. The search direction can be therefore defined as $\mathbf{g}^k = -\mathbf{H}^{-1}\nabla_{\mathbf{q}}(\mathbf{q})$. Several variants, approximating the Hessian matrix $\hat{\mathbf{H}}$, have been presented. For instance, the Gauss-Newton method allows replacing the Hessian Matrix $\mathbf{H}$ from its first order derivative by using $\hat{\mathbf{H}} = 2\nabla_{\mathbf{q}}^T\nabla_{\mathbf{q}}$.

- **Levenberg-Marquart:** As used by Ledesma-Carbayo et al. (2001), Levenberg-Marquart optimization scheme can be seen as a combination between Gauss-Newton and Gradient Descent methods. The search direction can therefore be written as $\mathbf{g^k} = -((\mathbf{H}^{-1}) + \delta\ \mathbf{I})\nabla_{\mathbf{q}}(\mathbf{q})$. The term δ allows automatically regulating the Gauss-Newton term (speed) and the Gradient Descent term (stability). When the estimated parameters are far from their optimum, the δ parameter increases and the optimization strategy can be seen as gradient descent strategy. Inversely, the coefficient δ is decreased when the parameters are close to their optimum and the optimization strategy can be seen as Gauss-Newton strategy.

3.2 Existing methods

From the previous section, we can observe that tracking or registration methods depend on different components, i.e. the matching criterion, the transformation model, the regularization term and the optimization strategy. In this section, we describe related work dedicated to motion estimation in ultrasound images. As ultrasound tracking methods are closely related to multi-modal registration, we also present different registration approaches that allow aligning ultrasound and pre-operative images.

3.2.1 US tracking methods

As stated by Sotiras et al. (2013), registration and tracking methods are often categorized into intensity-based and feature-based approaches (De Luca et al., 2015). In this section, we relate the different tracking approaches and we present a classification of tracking methods according to their main features in tables 3.1 and 3.2.

US intensity-based approaches

In this section, we present different tracking methods dedicated to the position estimation of 2D and 3D anatomical structures from intensity correspondence measurements. The following approaches are presented chronologically:

- **Yeung et al. (1998)** introduced a 2D US tracking method that follows muscle contraction motion in consecutive ultrasound images by exploiting the spatial temporal coherence of ultrasound speckle noise. Their approach retrieves deformable motions by using block matching model and SSD criterion. To constrain the displacement field, the authors proposed a regularization scheme based on elastic operator. Their method is validated through simulation, phantom and real data representing muscle contraction motion.

- **Pennec et al. (2001)** proposed a method that allows tracking brain deformations in 3D consecutive ultrasound images. This approach estimates the displacement field by minimizing the similarity energy from Local Cross-Correlation (LCC) criterion and regularizing the transformation from Gaussian filtering. The method has been tested from US images of a pig brain at a post-lethal status and from phantom studies.

- **Cohen and Dinstein (2002)** presented a 2D tracking method combining classical block matching model and ultrasound specific similarity measure. This criterion assumes that the US images are degraded by Raleigh distribution multiplicative noise based on maximum likelihood motion estimation schemes. The approach has therefore the advantage to be well suited to high speckle noise. The method was validated on six in-vivo ultrasound sequences of fetuses.

- **Basarab et al. (2008)** introduced an approach that allows estimating 2D motion of structures in order to compute ultrasound elastography. The proposed method aims at handling highly localized motions by introducing multiscale bilinear deformable block matching models as illustrated in Fig. 3.2. The displacement of each block is estimated by using cross-correlation similarity criterion. Such method has been validated on two ultrasound sequences of the thyroid gland.

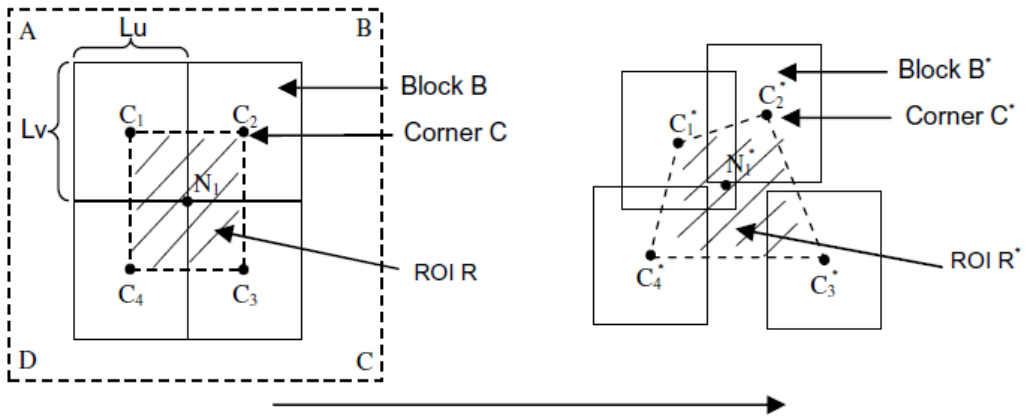


Fig. 3.2 Bilinear deformable block matching proposed by Basarab et al. (2008)

- **Elen et al. (2008)** developed a 3D tracking method in order to estimate the strain of the left ventricle from 3D US images. This approach is based on the estimation between free-form deformation and mutual information. In order to ensure tracking robustness, the displacement field is regularized by combining spatial smoothness and volume conservative constraint. The method was validated on in-vivo images acquired from 3 volunteers and one patient.

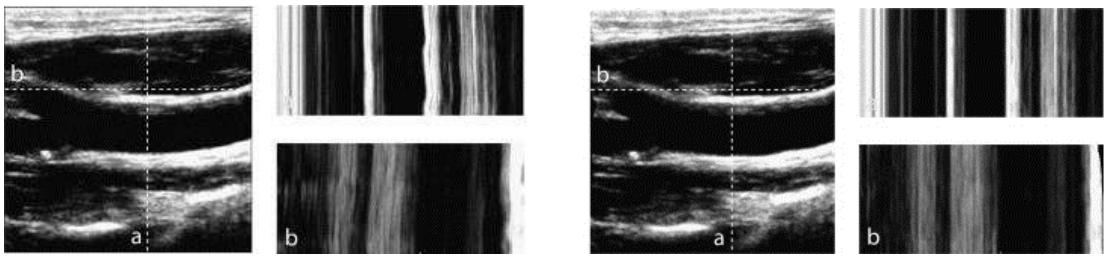
- **Wein et al. (2008b)** proposed an approach in order to compensate internal liver motion based on tracking sensors attached to both probe and skin. This method aims at registering 2D ultrasound slice to 3D volume acquired by sweeping the probe over the whole liver during a breath-hold. To do so, the image registration method combines affine model, local normalized cross correlation matching, and Powell-Brent optimization scheme. Their approach has been evaluated on liver sequences acquired from four volunteers.

- **Mukherjee et al. (2011)** provided a method that allows computing myocardial motion over 3D US sequence. This approach relies on the coarse to-fine displacement estimation of free-form transformation model obtained from SSD criterion. The validation was performed from 3D transesophageal echocardiography ultrasound images acquired on real patients.

- **Lee and Krupa (2011)** introduced a method that allows compensating of deformable soft tissue structures by controlling a 3D probe held by a robot arm. Such technique relies on the motion estimation of thin-plate spline model estimated from SSD criterion. Then, the robotic system compensates tissue deformation by extracting rigid motion

part from TPS motion model. This method has been validated by using two types of phantom.

- **Metz et al. (2011)** and **Vijayan et al. (2013)** proposed a new groupwise tracking strategy aligning all the US images of sequence at the same time. Each US image is taken into account in the computation of the cost function. In this approach, both spatial and temporal smoothness of the transformations are enforced by using a free-form B-spline deformation model. Some results of this method are illustrated in Fig. 3.3. In Vijayan et al. (2013), the method was validated on nine 3D ultrasound datasets of the liver of patients.



(a) Input image (b) Registration results

Fig. 3.3 Registration results obtained from group-wise registration method from Metz et al. (2011). The sub-figures (a) right top and right bottom represent the intensity of the initial image at the position of the dotted lines over the time. The sub-figures (b) right top and right bottom represent the intensity of the registered image at the position of the dotted lines over the time.

- **Bell et al. (2012)** introduced a 3D tracking approach based on rigid model where the displacements are computed from NCC matching criterion. Their study shows that volume rates as high as 8–12 Hz are needed for respiratory-dominated liver motion with 3D ultrasound speckle tracking in order to avoid speckle decorrelation.

- **Heyde et al. (2012)** presented an approach that estimates cardiac deformation within 2D US images. Such method relies on the improvement of free-form deformation model. As it can be seen in Fig. 3.4(left), free form deformation is generally represented by regular grid. The authors propose to update this formulation by integrating anatomical free-form deformation (AFFD) model where the basis functions are locally oriented along the left ventricle as shown in Fig. 3.4(right). This technique has the advantage to be very well-suited for cardiac motion. The displacements of the AFFD model are computed by optimizing the SSD criterion. The proposed method has been evaluated on five anesthetized sheeps.

- **Baumann et al. (2012)** presented a 3D ultrasound tracking method that allows retrieving prostate deformation from coarse-to-fine registration framework. Indeed, the method first allows retrieving the rigid motion from rigid model. Then, it allows computing the residual elastic motion from combination of ultrasound correlation-based criterion and elastic operator. The proposed matching criterion deals with local ultrasound intensity shift that may occur when the angle of US beam changes. These intensity shifts are compensated by using Gaussian convolution operators. The proposed approach was validated on a large set of 687 registered volumes obtained from 40 patients.

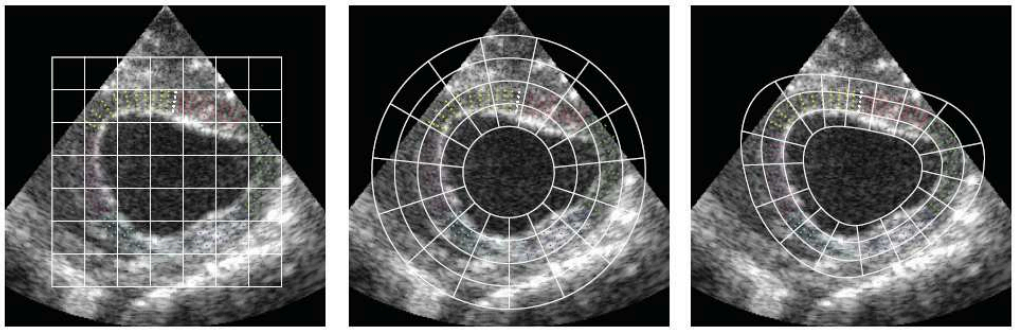


Fig. 3.4 Anatomical free-form deformation model proposed by Heyde et al. (2012)

- **De Luca et al. (2013)** introduced an approach that allows tracking liver vessels in long 2D US sequences. This method enhances the classical block matching method by introducing a strategy that allows adapting automatically block size. In addition, the proposed technique aims at improving the robustness by using a learning strategy that updates the reference image over the time. This method was validated on 9 sequences of the liver of volunteers under free breathing.

- **Masum et al. (2014)** presented an approach that allows measuring motion of the knee bones using tri-plane B-mode ultrasound. This method combines rigid model displacement and the Sum of Conditonal Variance criterion as this latter showed better results than Mutual Information, cross-cumulative residual entropy as demonstrated in Richa et al. (2011). No regularization term is required since rigid model constrains the displacement to be constant.

- **Banerjee et al. (2015)** proposed a new regularization framework inspired from game theory approach. This technique relies on the displacement estimation from classical block-matching methods. Then, the regularization framework removes false block matches from geometry and appearance constraints. The appearance constraint ensures

that the intensity error between matched block is not too high, while the geometric constraint allows removing blocks that show large displacement.

In table 3.1, the previous methods are classified according to their main components. We can observe that the majority of these approaches use mono-modal similarity metrics and therefore may provide incorrect results due to intensity variation produced by ultrasound shortcomings. To overcome this limitation, multi-modal similarity criteria have also been proposed such as Mutual Information (MI) (Elen et al., 2008), or Sum of Conditional Variance (SCV) (Masum et al., 2014). However, MI calculation is computationally expensive and is therefore not well suited for real-time tracking applications in 3D ultrasound images. The SCV has the advantage to be fast but it is not robust against local intensity variation that can be introduced by shadows. To cope with specific shortcomings of ultrasound imaging, US-specific similarity measures have also been evaluated by Baumann et al. (2012) and Cohen and Dinstein (2002). While these measures are well-suited to speckle noise and global intensity variation, they provide inaccurate results when US images are affected by shadows. To compensate the noise sensitivity of these criteria, most of the tracking methods are based on the usage of regularization terms. Several types of regularization can be identified such as spatio-temporal operators (Mukherjee et al., 2011), inverse consistency constraint (Baumann et al., 2012) and outlier rejection (Banerjee et al., 2015). Mechanical regularization has also been proposed despite its high computational cost (De Luca et al., 2015). However, such operator is not adapted for 2D tracking applications due to the presence of out-of-plane motions. From table 3.1, we can notice that these regularization terms can be combined in order to ensure better tracking results as proposed by Metz et al. (2011). Additionally, we can also notice that several approaches do not use any regularization since their transformation models sufficiently constrain the displacement field. However, such methods are not dedicated to the tracking of deformable targets. To improve the tracking accuracy, specific models have also been introduced since they are more adapted to targeted structures (Heyde et al., 2012) or US images (Basarab et al., 2008).

Method	2D/3D	Registration type			Manual Inter.	Pretreatment	Comp. Time
		Match. Crit.	Transformation	Regularization			
Lubke and Grozea (2014)	3D	SSD	Quadratic Poly.	Outlier rejection/Spatial	x	Histogram Equalization/Bilateral Filtering	61 ms
Touil et al. (2010)	2D	SAD	BM(Po)	Multiscale	x	x	x
Basarab et al. (2008)	2D	CC	BM(BD)	Multiscale	x	x	x
De Luca et al. (2013)	2D	CC(Learning)	BM (SA)	Temporal	x	x	100ms
Cohen and Dinstein (2002)	2D	US-ML	BM	x	x	x	x
Elen et al. (2008)	3D	MI	FFD	Spatial/Volume Conservative	x	x	10 min
Shekhar and Zagrodsky (2002)	3D	MI	Affine	x	x	Median Filtering/Intensity Quantization/TPVDI	1 min
Masum et al. (2014)	2D	SCV	Rigid	x	x	Intensity Quantization	x
Veronesi et al. (2005)	3D	SSD	Translational	x	x	x	1 min
Wein et al. (2008b)	2D	LNCC	Affine	Temporal	x	x	210 ms
Yeung et al. (1998)	2D	SSD	BM	Mechanical	x	x	x
Lee and Krupa (2011)	3D	SSD	TPS	x	x	x	100 ms
Heyde et al. (2012)	2D	SSD	C-FFD	Spatial	x	x	10,6 s
Ledesma-Carbayo et al. (2001)	2D	SSD	FFD	Forward backward direction	x	x	10 s
Pennec et al. (2001)	3D	SSD/LCC	Gaussian	Spatial/Temporal	x	x	10 min
Somphone et al. (2014)	3D	SSD	Gaussian	Fluid operator	x	Dark Filtering	100 ms
Mukherjee et al. (2011)	3D	SSD	FFD	Multiscale/Spatial/Temporal	x	x	x
Banerjee et al. (2015)	3D	NCC	BM	Outlier rejection (GeoAppCons)	x	x	125 ms
Metz et al. (2011)	2D	SSD (Av)	FFD	Inverse Consistency /Spatial/Temporal/GrpW.	x	x	90 min
Vijayan et al. (2013)	3D	SSD (Av)	FFD	Inverse Consistency/Spatial/Temporal/GrpW.	x	x	No
Marami et al. (2014)	2D	MIND	BM	Mechanical	x	x	380 ms
Bell et al. (2012)	3D	NCC	Translational	x	x	Median Filtering	x
Lediju et al. (2010)	3D	NCC	Translational	x	x	Median Filtering	x
Baumann et al. (2012)	3D	CC / SC	VF	Inverse Consistency/Mechanical	x	x	7 s
Angelova and Mihaylova (2010)	2D	MMPF	Deformable Contour	Spatial	x	Non-linear Gaussian filtering	4 min
Zhang et al. (2010)	2D	PDF	Deformable Contour	x	Surface Pts Ident.	x	3,4 ms
Rothlubbers et al. (2014)	3D	PDF	Rigid	x	x	x	372 ms
Mikic et al. (1998)	2D	SSD	Deformable Contour	Mechanical/Outlier rejection(VelFil)	x	x	6 s

Table 3.1 **Match. Crit.** Matching criterion, (SSD) Sum of Squared Difference, (SAD) Sum of Absolute Difference, (CC) Cross-Correlation, (SC) Shift-Correlation (US-ML) Ultrasound Specific Maximum Likelihood estimator, (MIND) Modality Independent Neighborhood Descriptor, (PDF) Probability density function, (StP) Surface to Points distance, (Learning) Learning-based method. (Av) Reference image intensity values corresponds to Average intensity value over time (SCCV) Sum of Confident Conditional Variance
Transformation (MPPF) Multiple Model Particle Filter, (BM) Block Matching, (Po) Polar, (BD) Bilinear Deformable, (SA) Scale adaptative, (VF) Variation framework, (Quad. Polyno.) Quadratic polynomial.
Regularization (Spatial/Temporal) Spatial/temporal smoothness, (Grpwise.) Groupwise, (VelF) Velocity filtering, (GeomCons/AppConstraint): Geometry and appearance constraint, (x) Not specified.
Manual Inter. Manual interaction during tracking task, (Surface Segm.) Surface segmentation, (Surface Pts Ident.) Surface Points Identification, (x) Not specified.
Pretreatement (TPVDI) Trilinear Partial Volume Distribution Interpolation, (Dark filtering) Points in the darkest regions are discarded. (Quadratic Poly.) Quadratric polynomial, (x) Not specified.
Comp. time Computation Time, (x) Not specified.

Feature-based methods

Feature-based methods aim at providing the spatial transformation between reference and current images by aligning key features, e.g. surfaces (Papademetris et al., 2002) or keypoints (Schneider et al., 2012). In this section, we introduce chronologically some approaches that rely on feature detection:

- **Papademetris et al. (2002)** have designed an approach that allows estimating motion of regional cardiac structure from three-dimensional image sequences. This method is based on shape tracking that is performed on interactively segmented images. In order to constrain the deformation field, the proposed algorithm used linear elastic operator. This method was validated on 3D in-vivo US images obtained from 4 dogs.

- **Comaniciu et al. (2004)** introduced an approach that allows tracking myocardial surface in 2D real-time images. In this method, the target is represented by b-spline curve and its position is estimated by using dynamic prediction obtained from Kalman filter. The motion estimation is performed by using optical flow algorithm on a sparse set of control points. To ensure robustness, a regularization term, combining adapted principal component analysis and uncertainty measure, is also introduced. The approach was validated on 32 ultrasound sequences acquired from patients.

- **Xiao et al. (2007)** presented a method for tracking anal canal in 2D ultrasound images. This technique is based on Mumford-Shah formulation that aims at combining an edge detection and an intensity-based approach. The displacement field is smoothed by using a regularization term based on curvature energies. The proposed method has been validated in a sequence of real anal ultrasound images evaluated by three experts.

- **Nascimento and Marques (2008)** also proposed an approach that allows tracking left ventricle of the heart from dynamic prediction. The main contribution consists in using multiple dynamic models in order to predict different abrupt motions of the heart (systole and diastole). The curve displacements are estimated by searching edge features along its normal. Outlier rejection process allows removing false edge features depending on their length and their distance from predicted contour. The approach was validated with 10 2D ultrasound sequences of the heart acquired from different patients.

- **Hu et al. (2008)** proposed a method that allows registering ultrasound images during prostate interventions. This approach has the advantage to provide robust tracking results by generating a patient-specific statistical motion model (SMM) of the prostate

from the combination of bio-mechanical simulation and boundary conditions such as the position of the TRUS probe/balloon and position and size of the pelvis relative to the prostate gland. However, its main drawback is related to the manual identification of landmark over each frame in order to perform registration. The method was validated on data obtained from 7 patients with prostate cancer.

- **Ni et al. (2008)** introduced a method initially dedicated to 3D ultrasound mosaicing. The proposed approach aims at combining intensity-based and feature-based matching. To do so, the initial registration is performed by using feature detection from Laplacian of Gaussian (LoG) filter. Then, the matching is locally refined by using block-matching that optimizes mutual information similarity measure. Furthermore, outlier rejection process allows removing block in shadowed regions. The method was validated from three ultrasound sequences obtained from phantom and real patient.

- **Angelova and Mihaylova (2010)** proposed a tracking method that allows providing lesion segmentation in 2D ultrasound sequence. The approach aims at evaluating the lesion contour from polyline model. The contour position is robustly estimated from dynamic prediction obtained from particle filter. The motion estimation is performed by an edge detection algorithm regularized by gradient norm. The approach was evaluated from several ultrasound sequences of breast cyst, thyroid and pancreas.

- **Schneider et al. (2012)** proposed a method that allows tracking heart motion in US volumes based on keypoint detection. To do so, potential features are extracted from 3D US volumes by using the LoG filter. Then, the descriptors of each feature are compared by minimizing Euclidean distance in order to find corresponding features between reference and current images. Outlier features are removed by using symmetric matching and Ransac algorithm. The method was validated with 34 ultrasound sequences of the mitral or aortic valve in a beating heart.

- **Häme et al. (2012)** introduced a method dedicated to the endocardium tracking in 3D US images. The proposed approach consists in estimating the motion of contour model represented by level-set formulation. The displacement estimation is performed by minimizing a cost function composed of several energy terms including image-based and curvature energy. The image-based energy allows following the displacement from edge detection. The second energy term regularizes the deformation from mean curvature constraint. The validation was performed with ten 3D sequences acquired from dogs.

- **Dietenbeck et al. (2014)** also proposed a tracking algorithm based on level-set formulation. The main contribution of this approach consists in finding the optimal contour position by adding different regularization constraints. To do so, the authors introduced cost function that includes motion estimation term, anti-collision term and shape prior term. The method was validated with 15 echocardiographic sequences acquired from patients.

In table 3.2, we propose to classify the previous methods according to their main characteristics. Contrary to the intensity-based approaches, these methods rely on the prior detection of features. The features are generally extracted from edge detection step (Häme et al., 2012) or Laplacian of Gaussian (LoG) maps generation (Ni et al., 2008). Once the detection is performed, the majority of these approaches are based on euclidean similarity measure in order to define the alignment between features. This criterion can define either the similarity between keypoint descriptors or the distance between sampled points of the contours. As the features are not always clearly visible, feature-based tracking approaches are often based on time prediction by using Kalman and Particle filters. Such regularization schemes allow providing better tracking results as it allows relying on prior knowledge. From this table, we can also observe that the choice of transformation model depends on the feature type. Generally, contour features are represented by either level-set or splines warping functions (Nascimento and Marques, 2008), while keypoints approaches propose to use rigid models (Schneider et al., 2012). To ensure accurate tracking, several hybrid methods have been proposed that combine intensity-based similarity criteria such as MI (Ni et al., 2008), SSD (Comaniciu et al., 2004), or monogenic phased matching (Dietenbeck et al., 2014).

Method	2D/3D	Registration type			Manual Inter.	Pretreatment	Comp. Time
		Match. Crit.	Transformation	Regularization			
Papademetris et al. (2002)	3D	Shape-Tracking	Gaussian	Mechanical	Surface Segm.	x	3 hours
Schneider et al. (2012)	3D	Euclidean(LoG)	Rigid	Ransac Filtering/ Forward Backward direction	x	Mask	170 ms
Ni et al. (2008)	3D	Euclidean(LoG)/MI	Rigid/BM	Ransac Filtering	x	Gaussian filter/ Shadow Removal	1 min
Yipeng Hu et al. (2011)	3D	StP	FEM	Statistical Motion Model (FEA)	Surface Pts Ident.	x	No
Comaniciu et al. (2004)	2D	SSD	Spline contour	SA-PCA/Kalman Filter(H)/ Outlier Rejection	Surface Segm. (PCA)	x	50 ms
Häme et al. (2012)	3D	Euclidean (E)	Level-set	Curvature Energy	x	Gaussian Filtering	x
Nascimento and Marques (2008)	2D	Euclidean (E)	Spline Contour	Kalman Filter (MMD) /Outlier rejection	x	x	x
Angelova and Mihaylova (2010)	2D	Euclidean (E)	Polyline	Particle Filter/Convexity constraint/Outlier Rejection	x	Non-Linear Gaussian Filter	6 min
Orderud et al. (2007)	3D	Euclidean (E)	Rigid/Spline Contour	Kalman Filter	x	x	40ms
Xiao et al. (2007)	2D	Euclidean/ GMM	Rigid/GMM	Curvature energy	x	x	x
Dietenbeck et al. (2014)	2D	MPM	Level-set	Anti-collision/ Shape	x	x	x

Table 3.2 **Match. Crit.** Matching criterion, (SSD) Sum of Squared Difference, (StP) Surface to Points distance, (E) Edge Detection, (MI) Mutual Information , (LoG) Laplacian of Gaussian, (GMM) Gaussian Mixture Model, (MPM) Monogenic Phase Matching
Transformation (Polyline) Line composed of multiple segments, (Level-set) Level-set formulation (FEM) Finite Element Model, (Shape) Prior shape term
Regularization (FEA) Finite Element Analysis, (H) Heteroscedastic noise (x) Not specified.
Manual Inter. Manual interaction during tracking task, (Surface Segm.) Surface segmentation, (x) Not specified.
Pretreatement (x) Not specified.
Comp. time Computation Time, (x) Not specified.

3.2.2 Multi-modal registration methods

In previous section, we described different tracking approaches that allow defining the target position over the time. In this section, we present some related work that allows registering pre-operative images of other modality to US imaging. These methods can be divided into two main categories, i.e. simulation-based and feature-based approaches. Simulation-based approaches consist in simulating a modality from another. Feature-based approaches allow mapping the different imaging modalities by extracting common features such as surface and vessels.

- **Roche et al. (2001)** introduced an approach for rigidly register real US and simulated US images obtained from MR volume. The simulation is performed by defining a function that estimate US intensities from MR intensities and MR gradient. The registration method is based on the optimization of bivariate correlation ratio (BCR) criterion that allows measuring alignment between real and simulated images. The proposed method aims at rejecting bad intensity matches from S-estimator. This approach was validated from data obtained from brain images of baby and patient.

- **Porter et al. (2001)** proposed an approach that registers MRI and US volumes from the extraction of common features. To do so, the liver vessels are detected from different images by using 3-D region growing segmentation algorithm. Then, the displacements of rigid model are estimated from maximization probability maps representing liver vessels. The method is evaluated on two set of in-vivo data.

- **Penney et al. (2004)** presented an approach that aligns 2D US images and MR volume of the liver. Their method is also based on vessels alignment from the US images where vessels are extracted by analyzing intensity level and intensity dips. Then, rigid motions are obtained by optimizing normalized cross-correlation criterion. The validation was performed between different data collected from five volunteers.

- **Lange et al. (2008)** introduced a method that registers 3D US and CT images of the liver based on thin-plate spline model. The displacements are computed by optimizing a cost function that combines intensity-based and feature-based terms. The intensity-based approach aims at finding the optimal transformation by using normalized gradient field, while feature-based is based on the alignment of manually annotated landmarks. The method was validated on data obtained from three patients.

- **Wein et al. (2008a)** introduced a method that allows registering real and simulated US images obtained from CT volume. The simulation is performed by mimicking

ultrasound ray-tracing scheme from acoustic impedance parameters estimated from CT. Then, the registration is based on combination between affine model and Linear Correlation of Linear Combination (LC^2) matching criterion. The LC^2 measure aims at evaluating the correlation between US intensities and different signals extracted from CT. The proposed approach was validated on ultrasound and CT data obtained from 19 patients.

- **Wein et al. (2010)** proposed a deformable method that provides dense-field deformable registration of CT and 3D ultrasound. It is also based on the simulation of US images from CT images. Nevertheless, this approach provides more robust results since it relies on the optimization of Local Cross-Correlation used in a variational formulation. The validation of the method was performed on 12 different patients with liver lesions.

- **Lee et al. (2011)** presented a non-rigid registration approach that aligns 3D ultrasound and CT images of the liver. This method allows retrieving the transformation by aligning surfaces and vessels from a cost function that combines mutual information criterion and Euclidean distance between surfaces. Furthermore, the displacement is regularized by using volume preservation constraint. The method was evaluated on 5 patients.

- **Kuklisova-Murgasova et al. (2012)** presented a method that registers MRI and US volume of fetal brain. This approach consists in simulating US images from pre-operative images. However, the simulation is based on a simplified model since the acoustic properties of the fetal skull can not be retrieved from MRI. The rigid displacements are computed by optimizing NCC matching criterion between images. The validation was performed on five datasets of fetal head.

- **Kadoury et al. (2012)** proposed a model-based rigid registration method that allows aligning 3DUS and MR images of the liver. This approach is based on liver surface extraction obtained from confidence based region algorithm. Then, surfaces are aligned by using a weighted Iterative Closest Point (ICP) algorithm that minimizes the distance between two set of points. To ensure robustness, a regularization term is introduced in order to penalize implausible transformation candidates that map probe positions further away from skin surface. The validation was performed on abdominal data obtained from four patients.

- **Nazem et al. (2013)** proposed a method that rigidly registers 2D ultrasound and CT images of the liver. This approach is based on the detection and the extraction of vessels and surface of the liver. The combination of ICP algorithm and Kalman Filter

allows extracting and predicting the motions of rigid model over each frame. The method was validated on a phantom study.

In table 3.3, we summarized the previous approaches with respect to their main components. Contrary to US tracking approaches, we can observe that most of these methods are not based on the usage of classical similarity measures such as sum of square difference, mutual information, or correlation ratio. Indeed, these criteria may not provide accurate results to register pre-operative and US images (Coupé et al., 2012). To cope with that issue, several approaches introduced novel similarity measures such as bivariate correlation ratio (Roche et al., 2001) or linear correlation of linear combination (Wein et al., 2008a). These similarity measures allow measuring the alignment between real US images and simulated images obtained from pre-operative data. They provide good performance with respect to shortcomings of US imaging. Another type of methods relies on identification of common features between US and pre-operatives images such as liver vessels and surfaces (Nazem et al., 2013). Despite the good performance of multi-modal registration approaches, we can also notice that they are not implemented in real-time due to their high computational cost (Lee et al., 2011). Furthermore, their transformation models are generally limited to rigid (Porter et al., 2001) and affine (Wein et al., 2008a) warping functions in order to strongly constrain the displacement field. As a consequence, the registration methods do not necessarily rely on the combination of regularization terms as their displacement field is already constrained by rigid or semi-rigid transformation models.

Method	2D/3D	Registration type			Manual Inter.	Pretreatment	Comp. Time
		Match. Crit.	Transformation	Regularization			
Roche et al. (2001)	3DUS/MR	BCR	Rigid	S-estimator	x	x	5 min
Kuklisova-Murgasova et al. (2012)	3DUS/MR	NCC	Rigid	x	x	x	x
Wein et al. (2008a)	3DUS/CT	LC^2	Affine	x	x	x	28 sec
Wein et al. (2010)	3DUS/CT	LCC	VF	x	x	x	10 sec
Coupé et al. (2012)	3DUS/MR	POF	x	x	Surface Segm.	x	2 min
Porter et al. (2001)	3DUS/MR	CC	Rigid	x	x	x	x
Penney et al. (2004)	2DUS/MR	NCC	Rigid	Gaussian Filtering probabilities	x	Gaussian Filter	x
Kadoury et al. (2012)	3DUS/MR	W-ICP	Rigid	Outlier transformation rejection	x	x	44s
Lee et al. (2011)	3DUS/CT	MI/Euclidean	Rigid/FFD	Incompressibility constraint	x	Anisotropic Filter	Several hours
Lange et al. (2008)	3DUS/CT	NGF/Euclidean	TPS	Elastic	x	x	10 min
Nazem et al. (2013)	2DUS/CT	ICP	Rigid	Undescent Kalman Filter	x	Anisotropic Filter	4,2 sec

Table 3.3 **Match. Crit.** Matching criterion, (BCR) Bivariate Correlation Ratio, (LC^2) Linear Correlation of Linear Combination, (NCC) Normalized Cross Correlation, (MI) Mutual Information, (ICP) Iterative Closest Point, (W) Weighted, (POF) Probability Objective Function, (LCC) Local cross-correlation
Transformation (FFD) Free-Form Deformation, (TPS) Thin-Plate Splines. (VF) Variational Framework
Regularization(x) Not specified.
Manual Inter. Manual interaction during tracking task, (Surface Segm.) Surface segmentation, (x) Not specified.
Pretreatement (x) Not specified.
Comp. time Computation Time, (x) Not specified.

3.3 Conclusion

In this chapter, we presented different approaches which aim at robustly estimating the position of anatomical structures in ultrasound images. Generally, these methods consist in improving different tracking components, e.g. matching criterion, transformation model, regularization operator and minimization strategy. Therefore, a first type of methods aims at introducing novel similarity criteria that are invariant to ultrasound imaging shortcomings such as speckle decorrelation and intensity variation (Cohen and Dinstein, 2002). To ensure robustness, another category of approaches is based on combination of different regularization terms in order to constrain the space of possible displacement field. These latter ensure spatio-temporal smoothness, target volume preservation, and inverse consistency of the transformation (Metz et al., 2011). Finally, several methods focused on novel transformation models in order to better represent the motions of targeted structures (Heyde et al., 2012).

Despite the number of contributions in this field, there are still some limitations that prevent determining the best components of the optimal tracking approach. A first issue is related to the lack of common validation dataset that makes the comparison between approaches difficult. Indeed, these latter are generally evaluated on different amounts of clinical images obtained from various imaging systems. A second limitation is associated to the dataset complexity. As stated previously, US images can be affected by different shortcomings including large shadows and speckle decorrelation. However, most of tracking and registration methods are evaluated on ideal conditions with US images that do not contain the previous shortcomings.

In the following chapters, we present different contributions in order to provide accurate tracking along 3D US sequences. We first propose a real-time approach that combines intensity-based method and physically-based model simulation. Then, we present a novel similarity criterion and tracking strategy that are specific to ultrasound imaging. These contributions have the advantage to be invariant to US imaging shortcomings. The proposed contributions are evaluated through simulated data, phantom data, and real-data. Furthermore, we compare our approach with respect to state-of-the-art techniques by evaluating our method on the 3D databases provided by MICCAI CLUST'14 and CLUST'15 challenges.

Chapter 4

Physically-based Deformable Tracking

In this chapter, we present a real-time approach that allows tracking deformable structures in 3D ultrasound sequences. Our method consists in obtaining the target displacements by combining dense motion estimation and mechanical model simulation. We also demonstrate that this approach can be adapted to raw data representing the voxel intensities along scanlines of the ultrasound probe.

As it can be seen from previous chapter, several methods have been proposed in order to track a target in 2D and 3D ultrasound images. To ensure robustness, multiple regularization terms have been evaluated such as incompressibility constraint, spatio-temporal smoothness and inverse consistency. Mechanical regularization has also been proposed as it allows removing non-physically plausible motions. For example, Yeung et al. (1998) and Marami et al. (2014) propose to combine intensity-based approach and mechanical simulation in order to follow target displacement in 2D sequence. However, such regularization may not be well-suited when out-of-plane motions are involved. To cope with that issue, Baumann et al. (2012) proposed an efficient 3D approach dedicated to prostate interventions by associating elastic and rigid registration from kinematic model of endorectal probe. Hu et al. (2012) proposed to improve the registration by using statistical motion model obtained from principal component analysis. However, their approach can not be used for real-time interventions as it relies on the manual identification of landmarks.

The chapter is structured as follows. In section 4.1, we provide a description of the proposed method. In section 4.2, we evaluate the method through phantom data and real data. Furthermore, we show the good performance of our method with respect to state-of-the-art techniques by testing on the 3D databases provided by MICCAI CLUST'14 and CLUST'15 challenges. In section 4.3, we further evaluate the method robustness regarding several parameters including model properties, the initial segmentation error, and the target velocity.

4.1 Method description

The objective of our approach is to track the motions of an anatomical target in a sequence of 3D ultrasound images. The first step of our method consists in generating a 3D tetrahedral mesh model associated to the target (section 4.1.1). This model is composed of a set of tetrahedral cells and a set of vertices. Once this model is defined, we estimate the target motions over the consecutive 3D images. For this purpose, the vertices displacements are computed by iteratively summing the internal displacements estimated from a mechanical component, and the external displacements computed from an intensity-based approach. The computational flow of the method is summarized in Fig. 4.1.

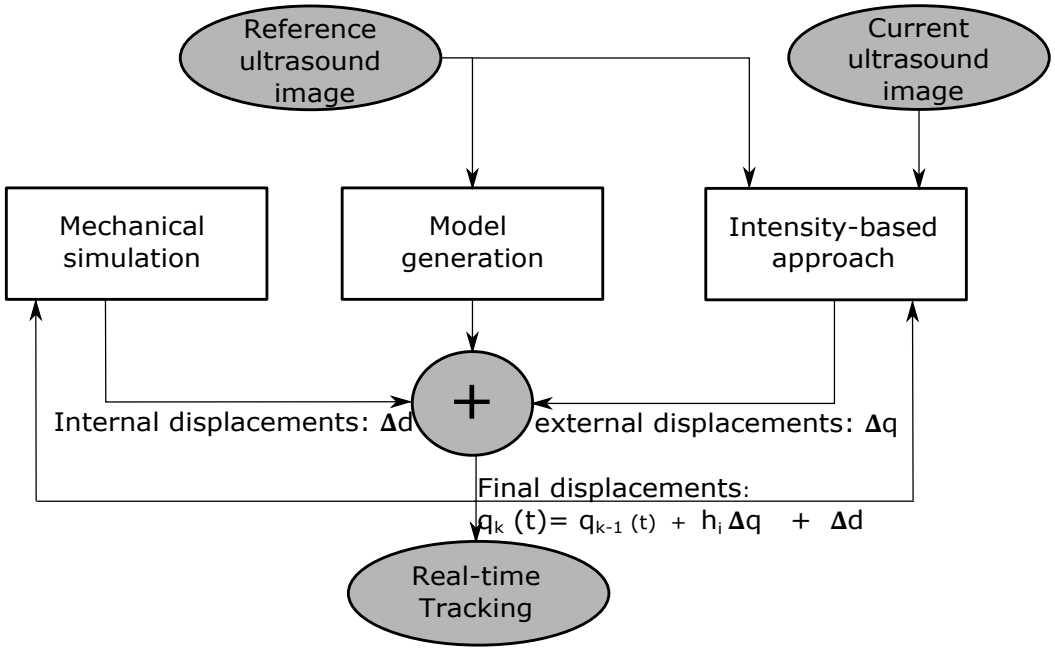


Fig. 4.1 Computational flow of the method. The different steps of the method are represented by white squares. The input and output data are characterized by grey ellipses.

As it can be seen from Fig. 4.1, the vertices positions $\mathbf{q}_k(t)$ of the model are computed by using the following equation:

$$\mathbf{q}_k(t) = \mathbf{q}_{k-1}(t) + h_i \Delta\mathbf{q} + \Delta\mathbf{d} \tag{4.1}$$

where $\Delta\mathbf{d}$ is the internal displacements obtained by integrating mechanical forces. $\Delta\mathbf{q}$ represents the external displacements estimated from the intensity-based approach. The computation of $\Delta\mathbf{d}$ and $\Delta\mathbf{q}$ are respectively described in sections 4.1.2 and 4.1.3. $\mathbf{q}_{k-1}(t)$ denotes the estimation of the previous vertice positions at time index t at iteration $k-1$ of

the optimization algorithm. h_i represents a coefficient that balances the contribution between image forces and internal mechanical forces. In section 4.1.4, we propose an adaptation that allows computing external displacements $\Delta\mathbf{q}$ from ultrasound raw data.

4.1.1 Model generation

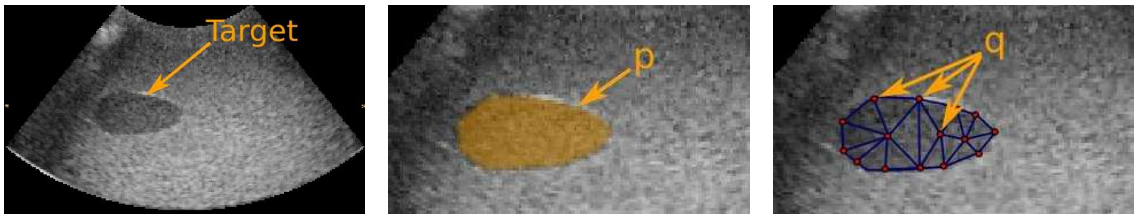
In 3D US images, an anatomical target can be represented by a continuous set of N_v voxels that is delimited by a visible border. In order to define a model representing the target, we first extract its shape in the initial 3D frame of the US sequence by performing segmentation. The segmentation is performed manually by segmenting the target within each 2D slice of the 3D volume. However, automatic segmentation methods can also be used (Chang et al., 2005). A smoothing step is performed on the 3D segmented surface in order to remove sharp edges and discontinuous shapes. Then, a corresponding fitted tetrahedral mesh model containing N_c vertices is defined on the segmented shape. Once the model is defined, we propose to use a piece-wise affine warp function that is parameterized from both the vertices positions and an affine interpolation using barycentric coordinates. In this way, we can relate all the voxel positions $\mathbf{p}$ with all the vertices $\mathbf{q}$ as follows:

$$\mathbf{p} = \mathbf{M}.\mathbf{q} \tag{4.2}$$

where $\mathbf{M}$ is a $(3 \cdot N_v) \times (3 \cdot N_c)$ constant matrix defining the set of barycentric coordinates. Each 3-line of $\mathbf{M}$ defines the set of barycentric coefficients regarding the x, y and z axis. $\mathbf{p}$ is a $(3 \cdot N_v)$ vector defining all the voxels positions, and $\mathbf{q}$ is a $(3 \cdot N_c)$ vector containing all the vertices positions. It is worth mentioning that $\mathbf{p}$ represents only the voxels in the mesh model. Thanks to Equation (4.2), we can update the positions of the target when the vertices of the model are displaced. To compensate the lack of smoothness as well as the poor estimation of vertices positions in US images, we combine a mechanical model to the estimation of displacement. The model generation step is illustrated in Fig. 4.2.

4.1.2 Estimation of internal displacements

Our approach combines a mechanical model based on mass-spring-damper system to the mesh previously described. Thus, the vertices displacements are constrained by linking each connected vertex pair with a spring and a damper ensuring physically-plausible and coherent displacements of the vertices. Furthermore, the mass-spring-damper system can be specifically characterized by setting a mass value to each vertex, together with elastic and damping coefficients on each spring depending on the soft-tissues homogeneities. From



(a) Original Image (b) Voxel positions (c) Vertices positions

Fig. 4.2 Illustration of the model generation step. (a) An anatomical region is identified in the first frame. (b-c) The associated voxel positions $\mathbf{p}$ and the vertices positions $\mathbf{q}$ are shown. For sake of clarity, tetrahedral cells are represented by triangular cells in this 2D illustration.

this model, we can compute the force $\mathbf{f}_{i,j} = [f_{x_{ij}}\ f_{y_{ij}}\ f_{z_{ij}}]^T$ applied on a vertex $\mathbf{q}_i$ from a neighbor vertex $\mathbf{q}_j$. This force can be expressed as follows:

$$\mathbf{f}_{i,j} = K_{ij}(d_{ij} - d_{ij}^{init})(\mathbf{q}_i - \mathbf{q}_j) + D_{ij}(\dot{\mathbf{q}}_i - \dot{\mathbf{q}}_j) \circ (\mathbf{q}_i - \mathbf{q}_j) \tag{4.3}$$

where d_{ij} and d_{ij}^{init} respectively represent the distances between the vertices $\mathbf{q}_i$ and $\mathbf{q}_j$ at their current positions and at their initial positions. The $\circ$ operator expresses the element-wise matrix product, K_{ij} is a scalar value denoting the stiffness of the spring that links the two vertices while D_{ij} is the damping coefficient value. By combining the previous equation for all the vertices, we can express the total amount of forces $\mathbf{f}_i$ exerted on each vertex $\mathbf{q}_i$ of the mesh model as follows:

$$\mathbf{f}_i = \sum_{r=0}^{N_i} \mathbf{f}_{i,r} + G_i \dot{\mathbf{q}}_i \tag{4.4}$$

N_i denotes the number of neighbors vertices connected to the vertex $\mathbf{q}_i$. G_i represents the velocity damping coefficient associated to the vertex $\mathbf{q}_i$. In order to obtain the internal displacements $\Delta\mathbf{d}$ associated to the mass-spring-damper system, we integrate the forces expressed in Eq. (4.4) with a semi-implicit Euler integration scheme such that:

$$\Delta\mathbf{d} = \Delta_t\ (\dot{\mathbf{q}} + \Delta_t\ \mathbf{f}^{int}) \tag{4.5}$$

where $\Delta\mathbf{d}$ is the internal displacement vector of the vertices. $\dot{\mathbf{q}}$ and $\mathbf{f}^{int} = \{\mathbf{f}_0\ \mathbf{f}_1 \cdots \mathbf{f}_{N_c}\}$ represent respectively the first time derivative of $\mathbf{q}$ and the internal forces vector. Δ_t denotes the integration time step. Such mechanical constraint can ensure the smoothness of the piece-wise affine warping function.

4.1.3 Estimation of external displacements

Let us recall that the main objective of our approach is to iteratively estimate both the external and internal displacements of the mesh. To compute the external displacements, we use an intensity-based method that consists in minimizing the cost function C expressed as follows:

$$\hat{\mathbf{q}} = \arg\min_{\mathbf{q}} C(\mathbf{q}) = \arg\min_{\mathbf{q}} E(I_t(\mathbf{p}(t)), I_{t_0}(\mathbf{p}(t_0))) \tag{4.6}$$

where I_t is a vector representing the US intensity of the volume acquired at time index t. I_{t_0} is a vector representing the US intensity of the initial volume. $\hat{\mathbf{q}}$ denotes the optimal positions of vertices. $\mathbf{p}(t_i)$ represents the voxel positions at time index t_i. E expresses the dissimilarity measure. In this chapter, we only evaluated the performance of Sum of Squared Differences (SSD) matching criterion. However, it is worth mentioning that other criteria are assessed in the following chapter such as Weighted Sum of Squared Differences (WSSD), Sum of Conditional Variance (SCV) and the proposed Sum of Confident Conditional Variance (SCCV). The SSD criterion allows defining the cost function C as follows:

$$C(\mathbf{q}) = \left(I_t(\mathbf{M}\mathbf{q}_t) - I_{t_0}(\mathbf{M}\mathbf{q}_{t_0})\right)^2 \tag{4.7}$$

where $\mathbf{q}_{t_i}$ denotes the vertices positions at time index t_i. Once the cost function is defined, the objective is to iteratively estimate the vertices displacements of the mesh model by minimizing the cost function C. To do so, we perform a Taylor expansion of the previous equation:

$$C(\mathbf{q}) \approx \left(\mathbf{J}\Delta\mathbf{q} + (I_t(\mathbf{M}\mathbf{q}_t^{k-1}) - I_{t_0}(\mathbf{M}\mathbf{q}_{t_0}))\right)^2 \tag{4.8}$$

where $\mathbf{q}_t^{k-1}$ represents the estimation of the parameters at time t at iteration $k-1$ of the optimization algorithm. $\Delta\mathbf{q}$ are the external vertices displacements. $\mathbf{J}$ denotes the Jacobian matrix associated to the cost function. This matrix relates the variation of the parameters $\Delta\mathbf{q}$ with the intensity variation of I_t. Each coefficient of the Jacobian matrix can be computed analytically as follows:

$$\mathbf{J}(u,v) = \nabla I_t(u + wN_v)\, \mathbf{M}(u+w, v) \quad \text{with } w = v \bmod 3 \tag{4.9}$$

where u and v denote respectively a specific row and a specific column of the Jacobian matrix. ∇I_t represents the gradient vector of the current 3D US image I_t. It can be expressed such that $\nabla I_t = \{\nabla I_{t_x} \nabla I_{t_y} \nabla I_{t_z}\}$ where $\nabla I_{t_x}, \nabla I_{t_y}, \nabla I_{t_z}$ are respectively the image gradients regarding the x, y and z axis obtained from 6×1 spatial derivative filters. In order to obtain the optimal displacements of the vertices, we chose to use a forward-additive steepest

gradient strategy. It is worth mentioning that more complex strategies like Gauss-Newton or Levenberg-Marquardt methods may provide better results (Baker and Matthews, 2004). However, they can not be implemented in real-time since they require the pseudo-inverse computation of large Jacobian matrix. From the steepest gradient method, the external displacement can be obtained as follows:

$$\Delta \mathbf{q} = -\alpha \mathbf{J}^T [I_t(\mathbf{M}\mathbf{q}_t^{k-1}) - I_{t_0}(\mathbf{M}\mathbf{q}_{t_0})] \tag{4.10}$$

where $\alpha > 0$ denotes the step size of the minimization strategy. $\mathbf{J}^T$ represents respectively the transpose of the matrices $\mathbf{J}$. The interested reader could refer to the appendix B that gives the full detail about the computation of external displacements.

4.1.4 Raw data adaptation

In the previous section, the external displacements were obtained from complete 3D reconstructed images. In this section, we introduce an adaptation of the method that allows dealing with ultrasound raw data. As illustrated in Fig. 4.3, raw data represent intensity of voxels located along ultrasound scanlines and does not involve interpolation. Touil et al. (2010) demonstrates that raw data use may help improving the tracking accuracy. Furthermore, its rectangular structure allows reducing the size of input information and better defining the limits of ultrasound images. Such characteristic is particularly well-suited for tracking partially visible targets that are not completely within the field of view. However, these benefits do not come without shortcomings since raw data is not geometrically correct when the transducer has curved shape. Furthermore, as illustrated in Fig. 4.3, raw data is not represented by a continuous set of voxels. Thus, to adapt the intensity-based approach, a conversion function ψ is required in order to rearrange the raw data. As shown in Fig. 4.3, ψ is the transformation that maps a point $\mathbf{p}'$ of the raw volume into the point $\mathbf{p}$ in the rearranged raw volume. From this conversion function, we can rewrite the Eq. (4.7) as follows:

$$C(\mathbf{q}_t^k) = \left(I_t^r(\psi(\mathbf{p'}_{im}(t))) - I_{t_0}^r(\psi(\mathbf{p'}_{im}(t_0)))\right)^2 \tag{4.11}$$

where $\mathbf{p'}_{im}$ represents the voxel positions of the target in raw data. I_t^r and $I_{t_0}^r$ denote the rearranged raw US intensity volume at time t and t_0. In the previous cost function, the points $\mathbf{p'}_{im}(t_0)$ are located along the scanlines. However, during tracking task, the intensity of their corresponding points $\mathbf{p'}_{im}(t)$ may be unknown since these points may be displaced between scanlines. To overcome this limitation, we used a method that interpolates the unknown intensity of a point $\mathbf{p}$ by using the inverse distance weighting (IDW) interpolation

of neighboring points. It can be expressed as follows:

$$\tilde{I}^r(\mathbf{p}) = \frac{\sum_{v=0}^{N_n} \gamma_v(\mathbf{p}) I(\mathbf{p}_v)}{\sum_{v=0}^{N_n} \gamma_v(\mathbf{p})} \tag{4.12}$$

where $\tilde{I}^r(\mathbf{p})$ represents the unknown intensity value associated to the point $\mathbf{p}$. $\mathbf{p}_v$ denotes the set of neighbors of $\mathbf{p}$ that are on ultrasound scanlines. N_n expresses the number of neighbors points. γ_v is the inverse distance between $\mathbf{p}$ and its neighbors $\mathbf{p}_v$. This distance is evaluated from rearranged raw data which presents a correct geometry. To adapt the method described in the previous section, we also need to design a new spatial derivative filter adapted to raw data. As previously stated, the image gradient of a voxel $\mathbf{p}$ is computed from 6×1 filters and takes into account the intensity value of nearest neighbors. Thus, by using the IDW interpolation technique, the image gradient of a voxel $\mathbf{p}'$ in raw data can be expressed as :

$$\nabla I(\psi(\mathbf{p}')) = \sum_{j=-3}^{3} w_j \tilde{I}^r(\mathbf{p}_j) \tag{4.13}$$

where $\mathbf{p}'$ represents a point in the raw volume. w_j depicts the filter coefficients, and $\tilde{I}^r(\mathbf{p}_j)$ represents the intensity of nearest neighbors of $\mathbf{p}$ within rearranged raw volume along the filter direction. By combining the previous equations, the computation of external displacements can be expressed as follows:

$$\Delta\mathbf{q} = -\alpha \mathbf{J}^T [I_t^r(\psi(\mathbf{M}\mathbf{q}_t^{k-1})) - I_{t_0}^r(\psi(\mathbf{M}\mathbf{q}_{t_0}))] \tag{4.14}$$

As stated previously, one main shortcoming of US imaging is its small field of view that may prevent visualizing the complete target. This issue may have a significant impact on the tracking accuracy. However, thanks to the rectangular geometry of the raw data, we can easily detect which part of the target is within the field of view. Then, these positions can be ignored by removing them from the previous cost function. The rest of the method remains unchanged.

4.2 Results

In this section, we first describe the implementation details by specifying the parameters values of the method and the software libraries we used to implement our algorithm. Then, two types of experiment were conducted. The first type is performed on phantom data and consists in demonstrating the relevance of the combination between mechanical model and

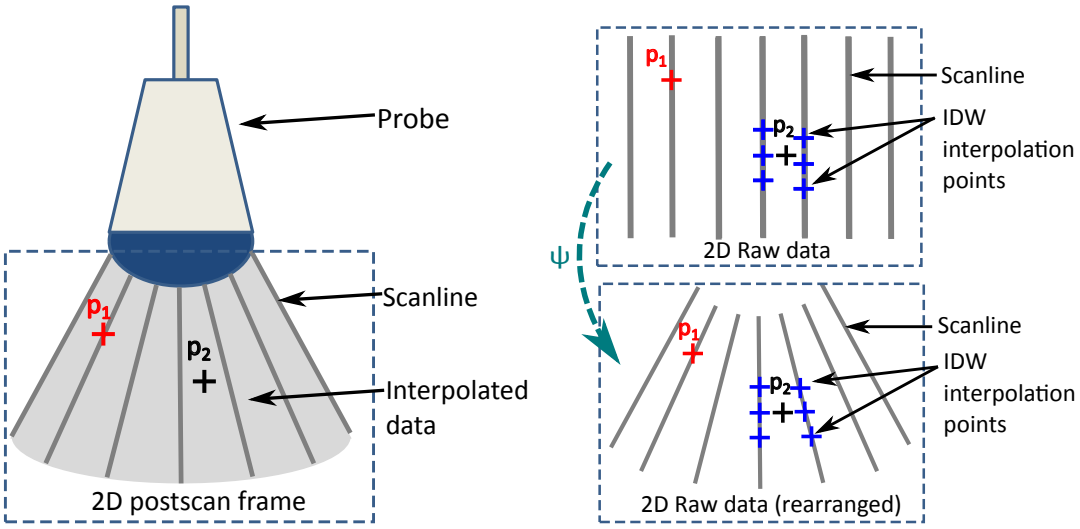


Fig. 4.3 Acquisition of raw data and postscan data. (left) Illustration of US image acquisition containing US scanlines and interpolated data. (right) Illustration of original rectangular raw data and rearranged raw data. ψ represents the conversion function.

intensity-based approach. We also propose to evaluate the adaptation that allows using the method on raw data. The second type of experiments aims at assessing the relevance against different state-of-the-art approaches. For this purpose, our method is tested on real-data provided by MICCAI CLUST'14 and CLUST'15 databases.

4.2.1 Implementation details

Our approach is implemented with C++/cuda code by using Cublas (Nickolls et al., 2008) and VTK (Schroeder et al., 2002) libraries. The code was executed on a Windows 7 machine with an Intel core i7-3840qm(2.80GHz) and achieves 350 ms for a tracking task between two consecutive 3D images. The initial segmentation and the tetrahedral mesh generation are performed respectively with ITK-SNAP (Yushkevich et al., 2006) and tetGen (Si, 2015) software. We set the elastic and damping parameters such that $K_{ij} = 3.0$ and $D_{ij} = 0.1$ for all the springs, along with $G_i = 2.7$ for all vertices. The step size of the steepest gradient method and the balancing coefficient have been respectively set to $\alpha = 2 \times 10^{-6}$ and $h_i = 1$. The number of iterations of the optimization algorithm has been set to 100 in order to ensure convergence. These parameters have been found empirically and are kept constant for the tracking tasks of all ultrasound sequences. In the section 4.3, we discuss the robustness of the approach regarding the properties of the model and its parameters.

4.2.2 Results on phantom data

As a first step of validation, we used the experimental setup illustrated in Fig. 4.4. The ultrasound images are acquired at frame rate of 0.8 vol/s from an Ultrasonix station (SonixTablet) with a motorized probe (4DC7-3/40). Each volume contains 44 2D images with an angular step of 1.4° representing a total field of view of 64.4°. The ultrasound volumes respectively consist of $128 \times 120 \times 44$ voxels. The voxel size represents a cube with 1.33 mm length sides. As illustrated in Fig. 4.4, the ultrasound probe is mounted on the end effector of a 6 degrees-of-freedom robot (Viper 650, Adept). The probe is located on the surface of a home-made phantom composed of gelatin and animal tissue. Thanks to this experimental setup, we can apply rigid motion and deformation to the phantom by displacing the probe regarding the x-y-z axis. In the following experiments, the validation of the approach is performed by using several metrics including barycenter displacement, intensity error, volume evolution of the model, and visual assessment. Furthermore, as the probe is mounted on the robot, we can compare our tracking results with the measurements of the robot displacements by using odometry.

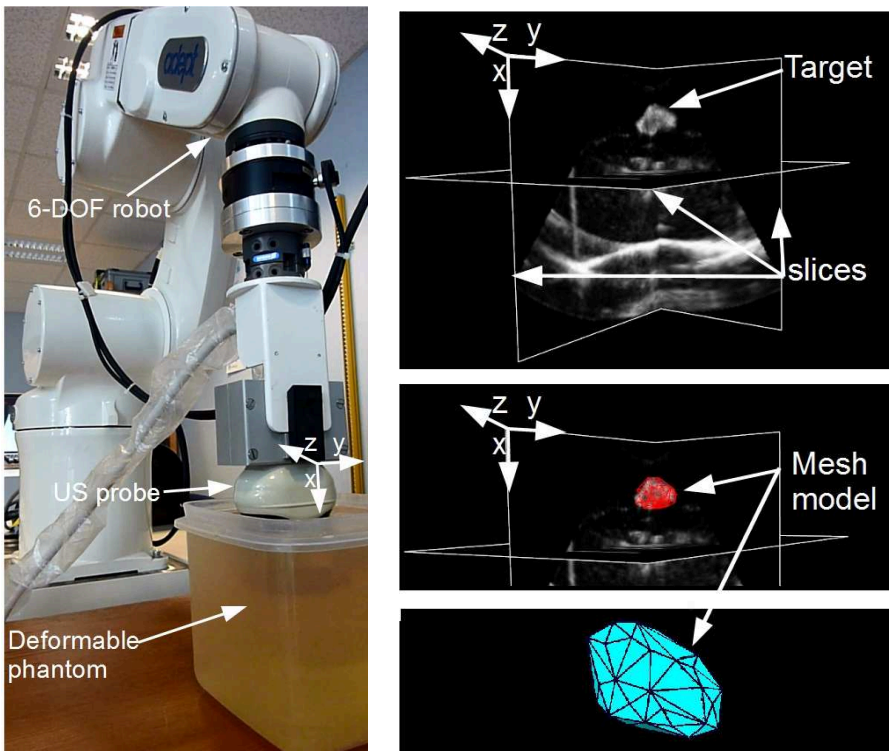
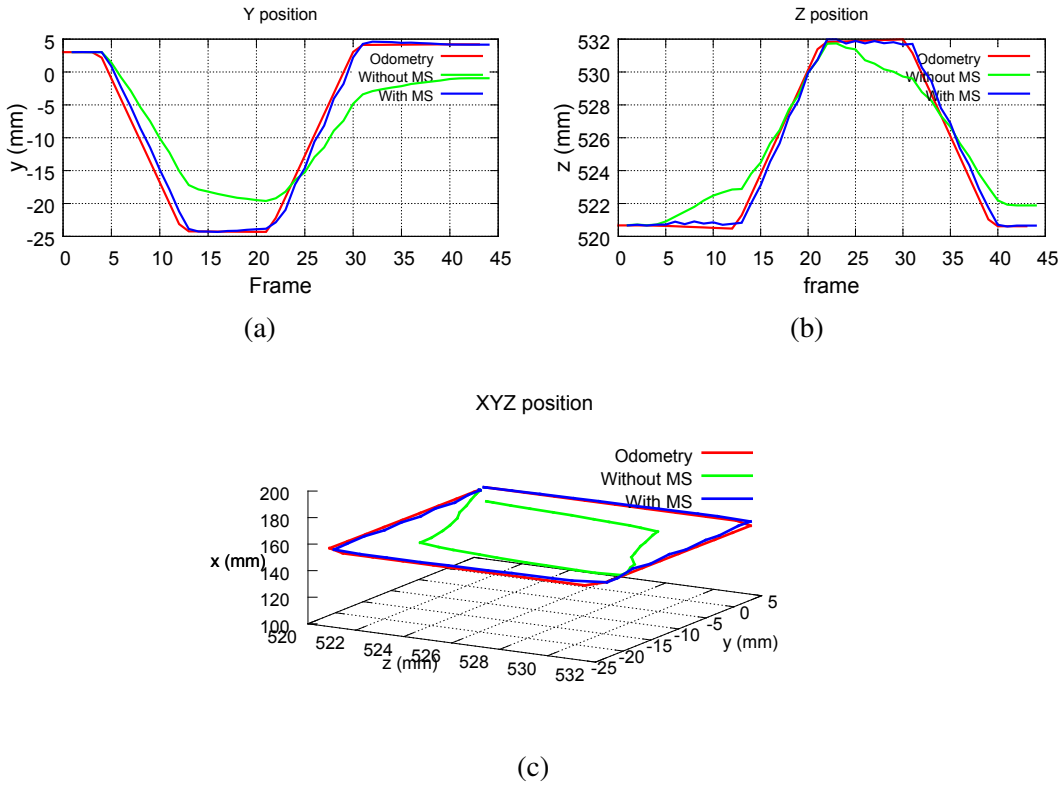


Fig. 4.4 (Left) Experiment setup. (Right) 3D slices of the ultrasound volume representing the target and the tetrahedral mesh model (148 cells, 60 vertices).

Phantom experiments with rigid motion

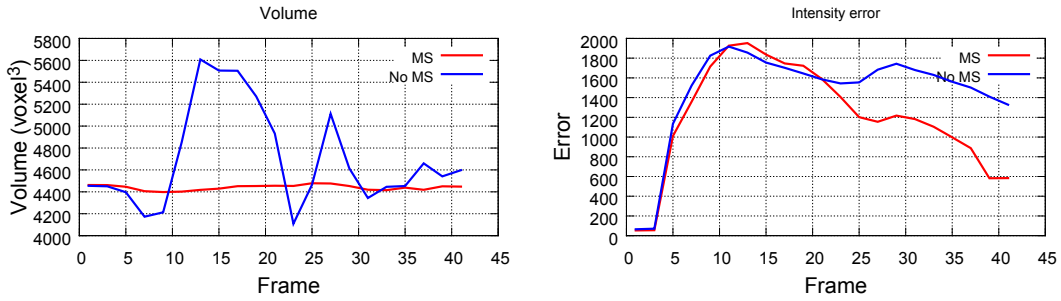
The first experiment consists in testing our approach with rigid motion. For this purpose, the probe is displaced along square trajectories regarding the y-z axis by using the robot. To demonstrate the relevance of the combination of piecewise affine approach and the mechanical model, we compare our method without the mechanical model by removing the $\Delta \mathbf{d}$ term of Eq. (4.1). In Fig. 4.5, the evolution of the barycenter position is shown during the tracking task. From this figure, we can observe that our method achieves a smaller mean tracking error (1.23 mm$\pm$0.03 mm) than the method without mass-spring system (5.25 mm$\pm$0.42 mm) regarding the odometry measurements. In Fig. 4.6, we show the evolution of the least square intensity error and the 3D mesh volume over each frame. We can notice that the method without mechanical component provides higher final intensity error and volume evolution that varies significantly. However, the target volume should remain constant since only rigid motion is applied. This can be explained by the presence of local minimum due to speckle noise perturbing the tracking. From the previous results, we can conclude that the mechanical regularization allows compensating the presence of local optimum since the mechanically-regularized method provides a smaller mean tracking error, smaller final intensity error and constant volume evolution. In Fig. 4.7, we illustrate some tracking results during a rigid motion including rotation motion by rotating the probe along the x-axis.



(a) (b)

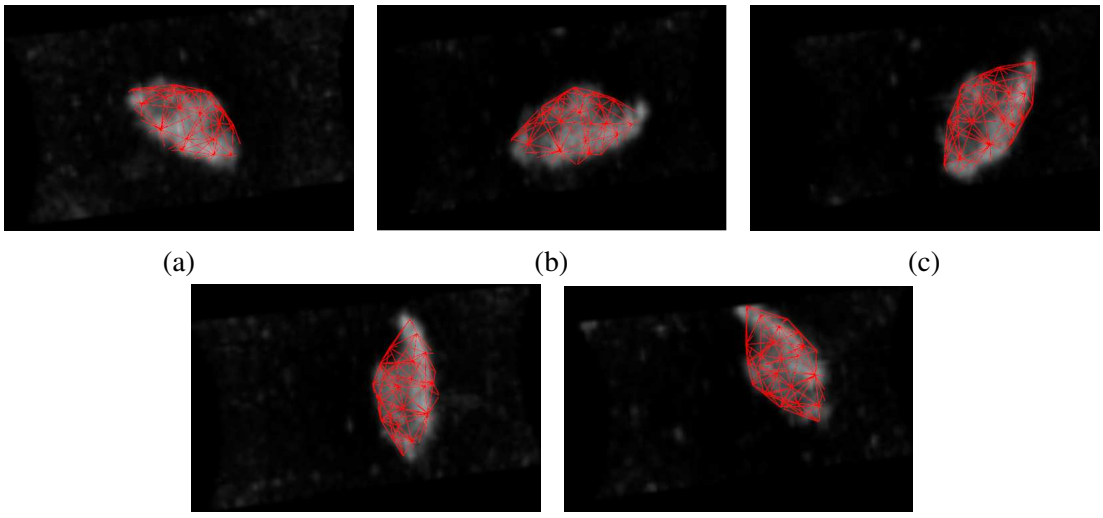
(c)

Fig. 4.5 Results obtained during lateral translation motion with and without physical model (Mass-Spring system (MS)). (a) Displacement of the target regarding the Y-axis over the time. (b) Displacement of the target regarding the Z-axis over the time. (c) 3D displacement of the target.



(a) (b)

Fig. 4.6 Results obtained during lateral translation motion with and without Mass-Spring system (MS). (a) Volume estimation of the target. (b) Least square intensity error.



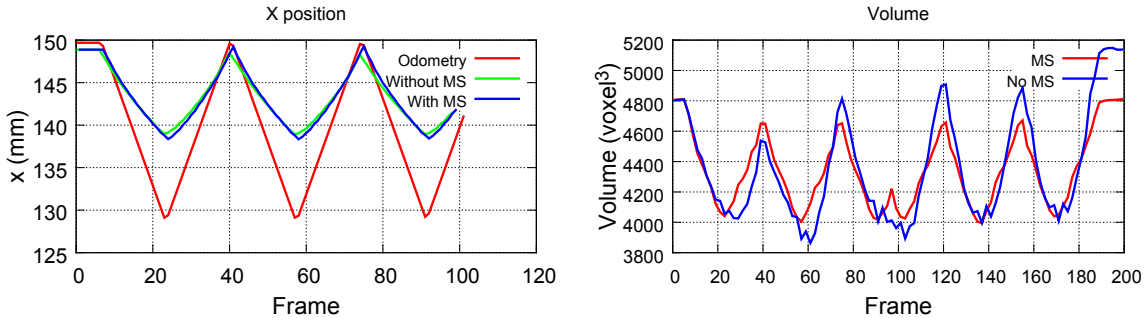
(a) (b) (c)

(d) (e)

Fig. 4.7 Example of the tracking task. Mesh model during tracking with rotation motion at frame 1 (a), at frame 16 (b), at frame 26 (c), at frame 36 (d), at frame 46 (e).

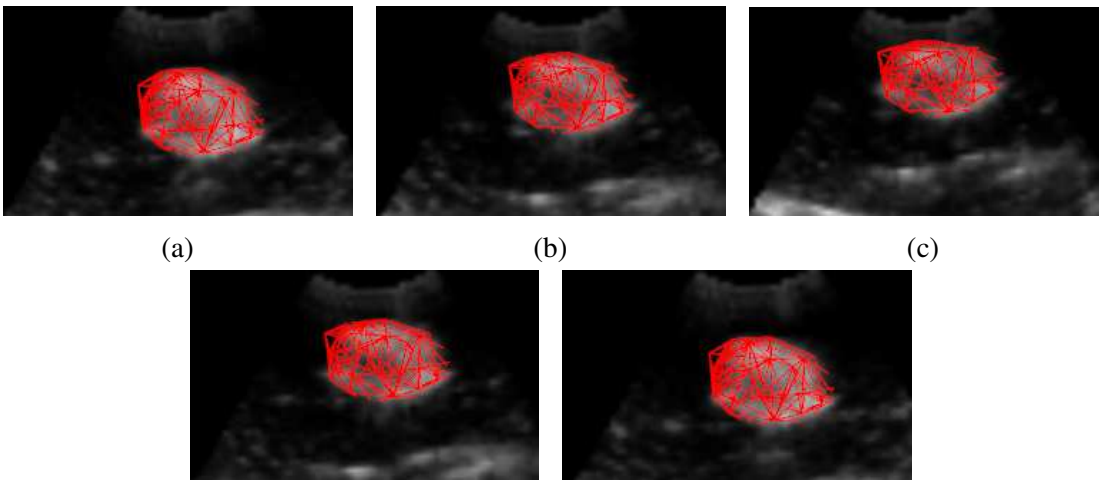
Experiment with deformation

In a second experiment, we evaluated our method during target deformation by applying a periodic motion to the probe along the x-axis that produces significant and periodic compression of the target. The estimated displacement of the target and the volume evolution of the 3D model during the tracking task are shown in Fig. 4.8. Contrary to the previous experiment, we can observe a large mean tracking error between the barycenter position estimated from odometry and from our tracking method. This can be explained by the fact that full motion can not be measured by odometry due to compression of both the target and neighborhood gelatin. To cope with this issue, we manually annotated a specific landmark of the target over each volume. The mean tracking error is computed by comparing the position of the annotated landmark with the warped position of the model over each volume. Thus, we can observe that both approaches provide approximately the same mean tracking error of 1.50 mm. However, in Fig. 4.8, we also notice that our method produces a steady periodic evolution of the volume varying with the probe displacement contrary to the method without mechanical component. From the previous results, we can see that our method allows avoiding local optimum by providing well-shaped model that fits the target surface over several frames, as well as steady volume evolution. The tracking results during deformation motions are illustrated in Fig. 4.9.



(a) (b)

Fig. 4.8 Results obtained during deformation motion. (Left) Displacement of the target regarding the X-axis over each frame. (Right) Volume estimation of the model over each frame.



(a) (b) (c)

(d) (e)

Fig. 4.9 Example of the tracking task. Mesh model during tracking with deformation motion at frame 1 (a), at frame 16 (b), at frame 24 (c) , at frame 32 (d), at frame 40 (e).

Experiment with raw data

In this experiment, we compare the method applied on ultrasound raw data and the original method applied on postscan data. The performance of the approaches is also evaluated when the target becomes partially visible during the sequence. To do so, we acquired a sequence where the robot followed a large square trajectory as well as an elevation motion which deformed the target. Fig. 4.10 shows the displacements of the model barycenter and the comparison with odometry measurement. We observed that the adaption improves the

tracking accuracy (1.94 mm±0.18 mm instead of 3.25 mm± 0.20 mm with postscan data) when the target goes out of the field of view. Fig. 4.11 shows an example of the tracking task on several frames showing a slice of target which is displaced to the right and goes out of the field of view. We noticed that, with the original method, our model does not have a plausible shape when the target is not within the field of view (see Fig. 4.11b). This can be also visualized in Fig. 4.10d because the volume of the model is abnormally reduced when the target goes out of the field of view at frame 20. The original method remains however robust since the model returns to its initial shape when the target becomes visible again during its return to the initial position as shown in Fig. 4.11c. It is also worth mentioning that both methods obtain similar results when the target is entirely visible. Therefore, in the following experiments, only the original method method will be evaluated.

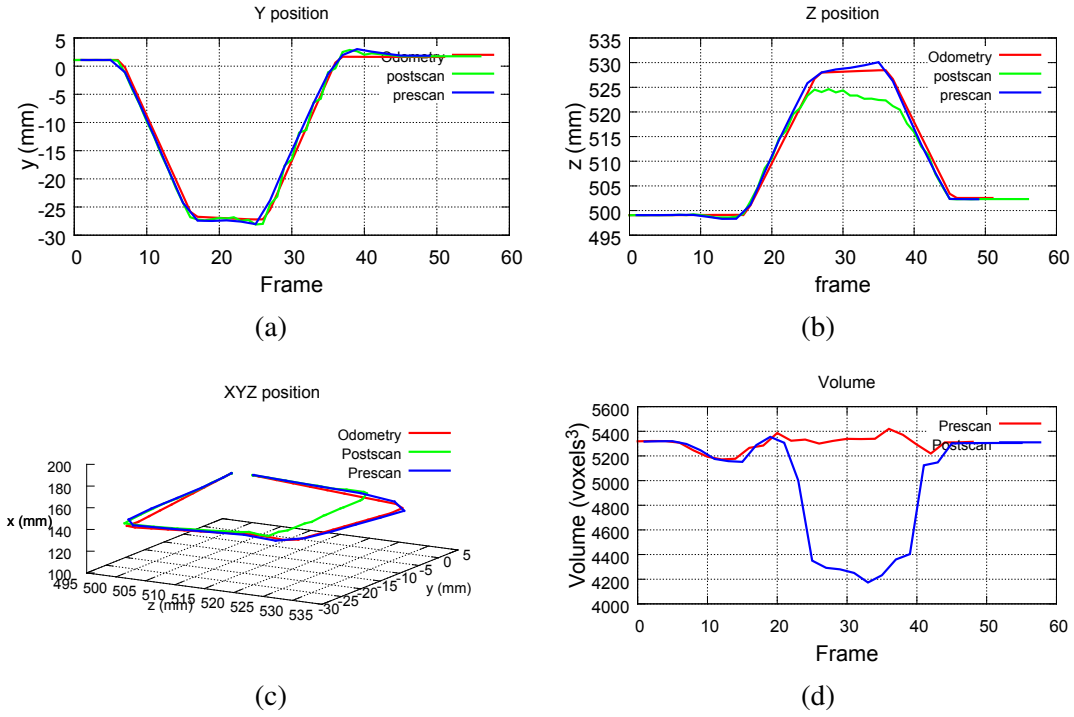
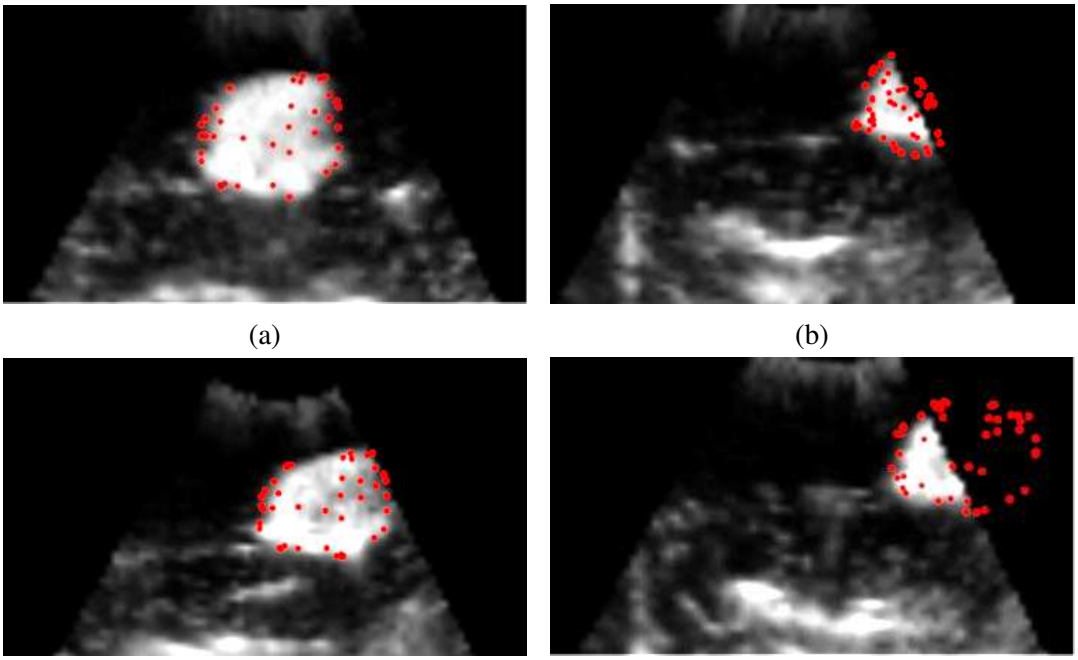


Fig. 4.10 Results obtained with partially visible target. (a) Displacement of the target regarding the Y-axis over the time. (b) Displacement of the target regarding the Z-axis over the time. (c) 3D displacement of the target. (d) Comparison of the volume estimation of the target with the prescan method and with the postscan method.



(a) (b)

(c) (d)

Fig. 4.11 Example of the tracking task with partially visible target experiment. The different images show a slice of the target regarding the Y-axis. Red points represent the intersections between the mesh model and the image slice. Tracking with postscan data (a) at frame 1, (b) at frame 34, (c) at frame 42. (d) Tracking with raw data at frame 34.

4.2.3 Results on real data

A second validation step is performed on real-data by using the 3D database provided by MICCAI CLUST'14-15 challenges containing ultrasound sequences of volunteers under free breathing. The main goal of these challenges is to compare different approaches for tracking anatomical landmarks in US sequences. The ground truth data is provided by using the manual annotations from three experts of target positions over each frame. Thus, a comparison can be performed between the ground truth landmark positions and the warped point positions (estimated from our model) over each frame. Our approach has been tested by tracking 34 different anatomical features acquired from 22 3D US sequences. The final tracking results on CLUST'14 and CLUST'15 databases are respectively reported in table 4.1 and 4.2. They are also presented on the challenge website[1,2].

Participants	**Mean**	**SD**	**95%**
Our method	1.62	2.19	4.81
Somphone et al. (2014)	2.55	2.46	7.98
Rothlubbers et al. (2014)	2.80	2.96	7.94
Lubke and Grozea (2014)	4.63	4.03	12.44

Table 4.1 CLUST'14 results of 3D point-landmark tracking expressed in millimeters. The first column of the table details the reference to each candidate method. The subsequent columns represent respectively the mean error, the standard deviation, and the 95th percentile expressed in millimeters for each approach.

Participants	**Mean**	**SD**	**95%**
Our method	1.74	0.92	3.65
Banerjee et al. (2015)	1.80	1.64	3.41

Table 4.2 CLUST'15 results of 3D point-landmark tracking expressed in millimeters. The first column of the table details the reference to each candidate method. The subsequent columns represent respectively the mean error, the standard deviation, and the 95th percentile expressed in millimeters for each approach.

From the previous results, we can observe that our method provides the smaller mean tracking error for the different databases. We can also notice that our tracking error is different between table 4.1 and 4.2 since the evaluation is not performed on the same database. The detailed results for all target tracking tasks are presented in table 4.3. As it can be seen in this

[1]CLUST'14: http://clust.ethz.ch/clust2014.html
[2]CLUST'15: http://clust.ethz.ch/results.html

table, the specific annotated landmarks are generally vein bifurcations since they are clearly visible within ultrasound images. We can notice that our approach shows good performance since the mean tracking error is low (<3 mm) for most of the ultrasound sequences except for three ultrasound sequences (SMT-04_1, EMC-03_1, EMC-07-3_1). We also observe that our method may produce inaccurate results when the 3D mesh model goes out of the field of view (SMT-04_01). However, this issue can be addressed by using either binary mask or rectangular raw data as presented previously. Inaccurate results can also be observed when the target follows high deformation ($> 50\%$) regarding the provided elastic parameters (EMC-03_01). This problem can be solved by using specific elastic parameters for each target. Furthermore, we also notice that the error can be higher due to strong motions between consecutive ultrasound frames (EMC-07-3_1). This issue can be tackled by using another ultrasound probe with higher frame rate. In the section 4.3, we detail the method robustness regarding the maximum amplitude of motion and deformation of the target.

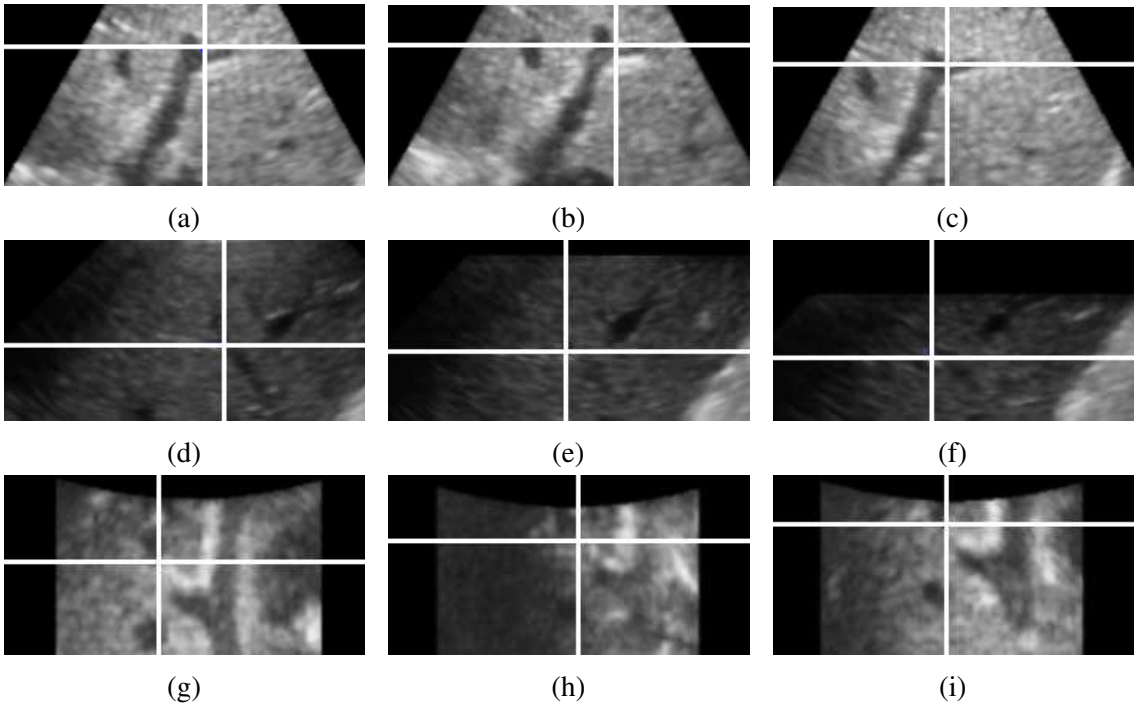


Fig. 4.12 Examples of the tracking task on several sequences with our approach. (a-b-c) Tracking of landmark representing hepatic vein bifurcation. The white cross represents the point position in Y-slice at frame index (a) 00, (b) 08, (c) 12. (d-e-f) Tracking of landmark representing vein birfucation in another 3D US sequence at frame index (d) 00, (e) 05 , (f) 12. (g-h-i) Tracking of landmark representing portal vein birfucation at frame index (g) 00, (h) 23, (i) 59.

Sequence	Min	MTE	Max	SD	95%	Landmark	Nb Vert.	Nb. Cells
EMC-02_1	0,24	0,65	0,88	0,26	0,88	H-VB	47	134
EMC-02_2	0,72	1,47	2,33	0,62	2,33	H-VB	40	107
EMC-02_3	0,66	1,07	1,38	0,27	1,38	H-VB	44	131
EMC-02_4	0,78	1,44	2,48	0,71	2,48	H-VB	65	152
EMC-03_1	0,59	3,37	5,73	1,71	5,67	H-VB	37	101
EMC-04_1	0,57	0,92	1,92	0,33	1,76	VB	26	63
EMC-05_1	0,21	0,55	1,13	0,24	1,07	x	39	107
ICR-01_1	0,43	1,52	2,99	0,85	2,93	x	30	73
ICR-02_1	0,25	2,73	4,09	1,08	3,97	x	46	140
SMT-02_1	0,48	1,06	10,15	0,98	1,3	H-VB	47	123
SMT-02_2	0,6	1,27	2,1	0,3	1,77	RB-H-VB	62	30
SMT-02_3	1,3	2,24	3,02	0,43	2,91	P-VB	27	53
SMT-03_1	0,61	1,3	2,29	0,46	2,03	Nephron	38	80
SMT-03_2	0,36	1,2	2,19	0,41	1,92	H-VB	41	100
SMT-04_1	0,65	9,28	14,43	4,16	13,77	KA	45	118
SMT-05_1	0,55	2,21	6,22	1,59	6,06	P-VB	30	79
SMT-05_2	0,15	0,62	1,58	0,24	1,07	H-VB	30	71
SMT-06_1	0,46	0,93	1,44	0,2	1,29	H-BDB	35	96
SMT-06_2	0,51	0,98	1,43	0,21	1,34	AB	30	83
SMT-06_3	0,43	1,12	2,17	0,41	2,01	AB	48	120
SMT-07_1	0,7	1,15	1,81	0,23	1,51	AB	30	65
SMT-07_2	0,77	1,28	1,84	0,27	1,76	VB	23	59
SMT-08_1	0,17	0,81	1,3	0,23	1,21	AB	33	83
SMT-08_2	0,19	0,64	1,11	0,17	0,89	VB	28	81
SMT-08_3	0,59	1,45	2,48	0,44	2,24	VB	48	143
SMT-09_1	0,13	0,77	1,76	0,3	1,41	VB	35	90
SMT-09_2	0,21	0,62	1,1	0,19	0,93	VB	36	99
SMT-09_3	0,17	1,39	2,26	0,46	2,18	VB	30	71
EMC-06-1_1	0,71	1,54	2,58	0,46	2,58	VB	70	198
EMC-06-2_1	0,95	2,95	4,45	1,16	4,45	VB	49	153
EMC-06-3_1	1	1,47	2,76	0,52	2,76	VB	64	184
EMC-07-1_1	0,73	2,66	5,92	1,69	5,92	VB	50	140
EMC-07-2_1	0,1	1,51	2,84	0,83	2,84	VB	67	187
EMC-07-3_1	2,59	3,54	4,48	0,77	4,48	VB	67	187

Table 4.3 Detailed tracking error results for each tracking task. The error results are obtained-from Euclidean distance and are expressed in millimeters. (Sequence) Name of the sequence (Min) Minimum tracking error. (Max) Maximum tracking error. (MTE) Mean tracking error. (SD) Standard deviation. (95%) 95th percentile of error. (VB) Vein bifurcation, (H-) Hepatic, (P-) Portal, (KA) Kidney Artery, (AB) Artery Bifurcation, (BDB) Bile Duct bifurcation.

4.3 Robustness evaluation

In this section, we evaluate the method robustness regarding several parameters including the gain h_i, the initial segmentation error, the mesh cell number and the motion amplitude. Finally, we assess the benefit of the mechanical regularization term. To evaluate the sensitivity regarding these parameters, we acquired three sequences detailed in table 4.4. The ground truth is obtained from annotations of landmark positions over each 3D volume from one expert. The tracking error is computed by comparing the position of the annotated landmark with the warped position of the model over each volume. To evaluate the accuracy of the landmark definition, we measured the inter-variability from the annotations of three observers along 20% of the total number of 3D frames.

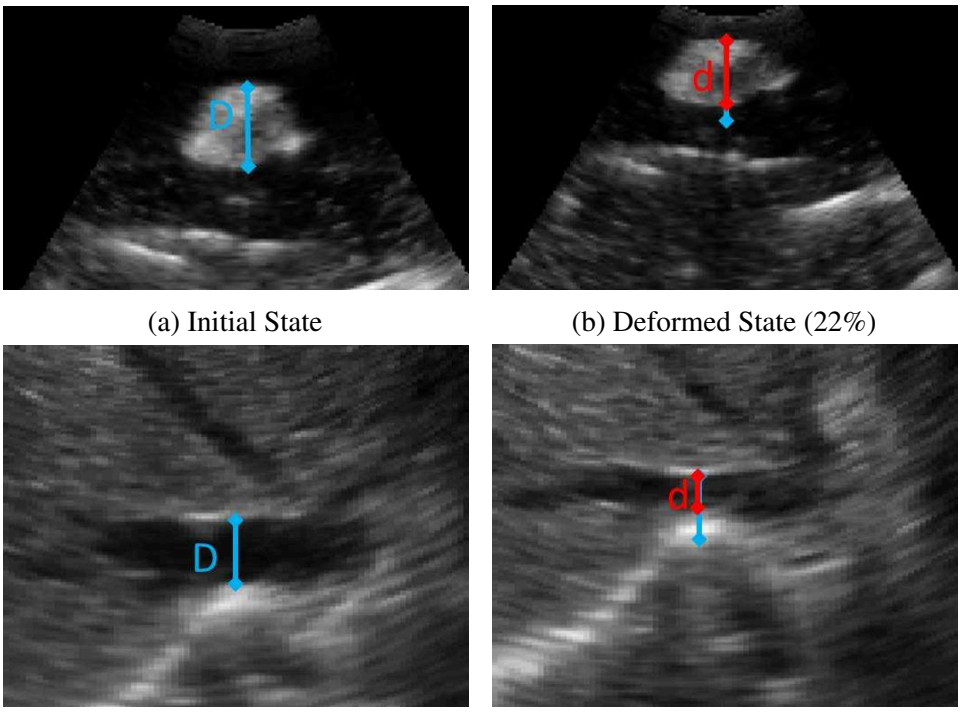
Sequence	Deformation	Target Dim. (mm)	Int. Displ. (mm)	Nb. Frame	Inter-var (mm)
PHA_5	Rigid	20 x 30 x 32	6.5	40	1.6 ± 1.0
PHA_6	22%	20 x 30 x 32	1.5	42	1.7 ± 0.8
PHA_7	49%	23 x 8 x 11	4.5	20	2.0 ± 0.7

Table 4.4 Details of sequences. (Sequence) Name of the sequence. (Deformation) Deformability of the target within the sequence. (Target Dim.) Target dimension expressed in millimeters. (Int. Displ.) Maximum target displacement between consecutive volumes expressed in millimeters. (Nb Frames) Number of frames in the sequence. (Inter-var) Inter-variability of the landmark annotation expressed in millimeters.

In the previous table, the deformation measurement is obtained by measuring the length between two landmarks of the target at the initial state and at the deformation state as follows:

$$M_{def} = \frac{d - D}{D} \tag{4.15}$$

where M_{def} denotes the target deformability that may be represented by a percentage value. D and d represent the initial and current length between the two landmarks. It is worth mentioning that the compressibility measure was not used since the volume of some targets can not be precisely retrieved due to inaccurate segmentation. In Fig. 4.13, we illustrate the deformability measurements of different targets.



(a) Initial State (b) Deformed State (22%)

(c) Initial State (d) Deformed State (49%)

Fig. 4.13 Example of deformable targets. (a-b) During the second experiment. (c-d) During the third experiment.

4.3.1 Mechanical parameter

Among the method parameters, the coefficient h_i is critical since it amplifies the contribution of external displacements. To evaluate the sensitivity of this parameter regarding the target deformability, we evaluate on different sequences the mean tracking error (MTE) of a specific landmark regarding the parameter h_i. In these experiments, the range of h_i values varies between 0.01 and 100 since it allows showing the limits of the method regarding this parameter. In Figs. 4.14, 4.15 and 4.16, we show respectively the evolution of the mean tracking error regarding h_i parameter value on sequences PHA_5, PHA_6 and PHA_7. Fig. 4.14 shows the tracking error of rigid experiment with respect to the value of parameter h_i. From this figure, we observe that the minimum tracking error value is obtained when $1 < h_i < 40$. The mean tracking error becomes larger when $h_i > 40$ due to the image noise sensitivity as the internal force contribution becomes small. We can also observe that the error is large when $h_i < 1$ due to the low external force contribution regarding the target displacement.

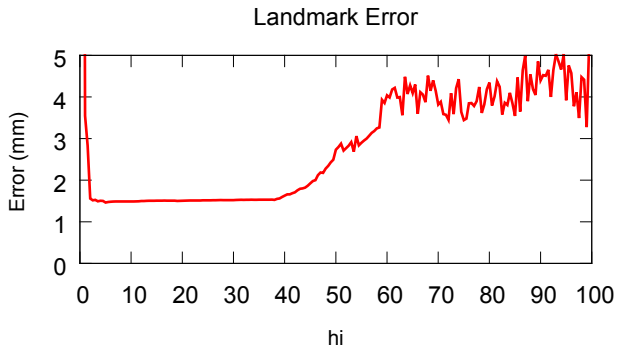


Fig. 4.14 Rigid tracking experiment: Evolution of MTE regarding parameter h_i.

In Fig. 4.15, the same experiment is performed on sequence whose target undergoes 22% of deformation. As it can be seen in the figure, the minimum tracking error value is obtained when $0.5 < h_i < 45$. The mean tracking error becomes larger when $h_i > 45$ due to the noise sensitivity that appears when the internal force contribution is too small. In the last experiment, we evaluate the sensitivity of h_i parameter on sequence that contains a highly deformable target (49%). Fig. 4.16 shows the evolution of the mean tracking error regarding the h_i value. As it can be seen in this figure, the minimum error ($< 2mm$) is obtained when $13 < h_i < 45$. Contrary to the other experiments, the range of h_i ensuring robust tracking is smaller due to the high deformation of the target.

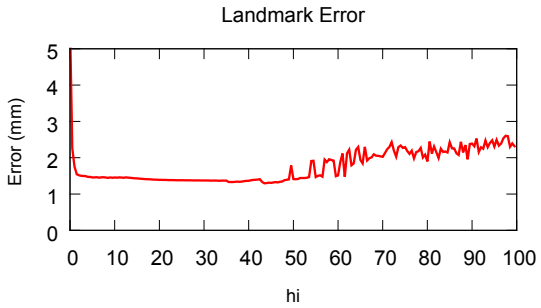


Fig. 4.15 Tracking experiment with deformable (22%) target: Evolution of MTE regarding parameter h_i

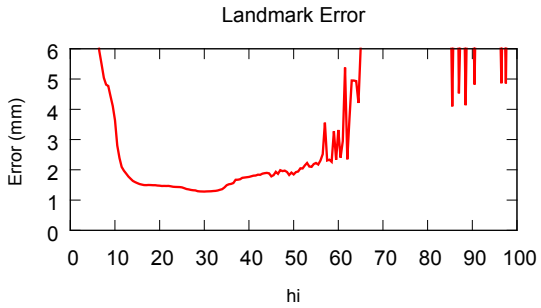


Fig. 4.16 Tracking experiment with deformable (49%) target: Evolution of the mean tracking error regarding parameter

As it can be seen from previous experiments, the tracking accuracy depends on the value of parameter h_i. Furthermore, we can also observe the range of possible values for h_i is larger when the deformation is limited (<22%). However, when the deformation is higher, the value of parameter h_i needs to be adjusted in order to emphasize external force contribution. In order to reduce the method's sensitivity regarding high deformation, an

interesting perspective would be to automatically update the parameter h_i according to the target elasticity that can be obtained from elastography images.

4.3.2 Initial segmentation

To illustrate the sensitivity of our approach regarding initial segmentation, we performed several experiments that consist in tracking same targets by using different segmentations. This evaluation is performed on the first two ultrasound sequences presented in this table 4.4 (PHA_5 and PHA_6). Contrary to previous experiments, we set $h_i = 1$ and we used different segmented models obtained from bad segmentations around the target. The segmentation error is measured by computing the Hausdorff distance between the initial (good) segmentation and the others. Thus, the higher is the Hausdorff distance, the worse is the segmentation. Fig. 4.17 shows respectively the MTE evolution regarding the Hausdorff distance on two sequences.

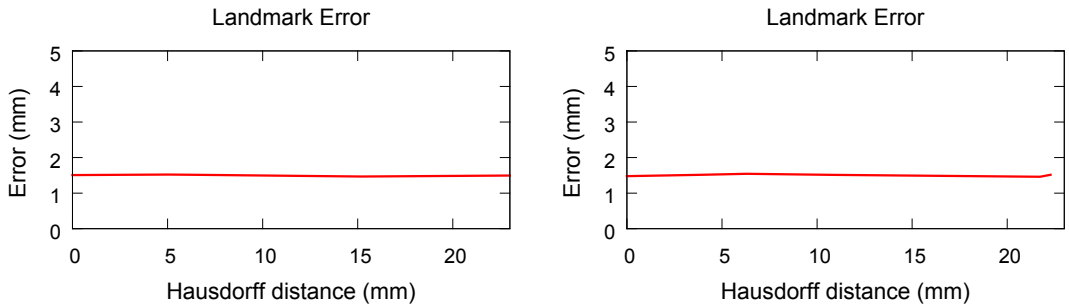


Fig. 4.17 Evolution of the mean tracking error regarding the segmentation error (Left) Rigid target. (Right) Deformable target (22% of deformation)

As it can be seen from Fig. 4.17, the tracking is not really sensitive to segmentation error even if it exceeds 2 cm. It is worth mentioning that this value is relatively important compared to the target size of 20 mm x 30 mm x 32 mm. Therefore, future work should consist in tracking objects that are automatically segmented from 3D US images by using method proposed by Chang et al. (2005) or Barbosa et al. (2014). Such methods may offer average segmentation errors up to 2.29 and 2.26 mm.

4.3.3 Mesh quality

To measure the method's sensitivity regarding mesh quality, we performed different experiments that consist in tracking targets by using mesh models with different number of both vertices and tetrahedral cells. This evaluation is performed on the two first sequences detailed

in table 4.4. In Fig. 4.18, we show the mean tracking error (y-axis) regarding the number of cells (x-axis) for rigid and non-rigid targets.

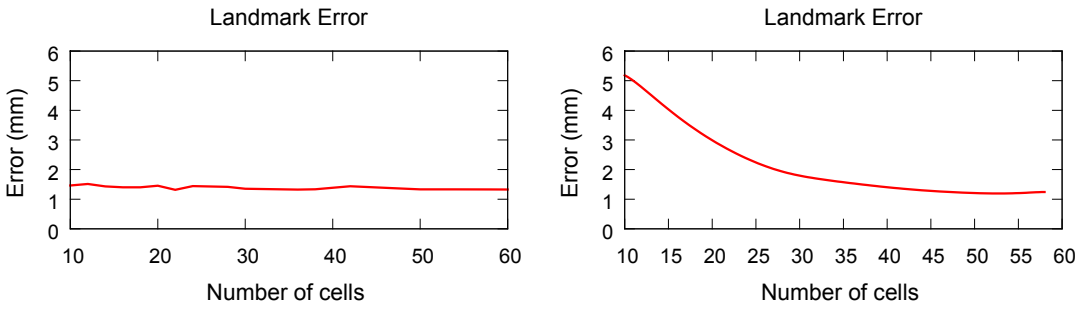


Fig. 4.18 Evolution of the mean tracking error regarding the number of cells of the model (Left) Rigid target. (Right) Deformable target (22% of deformation)

In Fig. 4.18, we can observe that the mean tracking error remains almost constant when the tracked object motion is rigid even if the cell number is low. This can be explained by the fact that all voxels of the target undergoes the same rigid displacement. In the non-rigid experiment showed in Fig. 4.18, we can observe that the error increases when the mesh is reduced. This can be explained by the fact that a sufficient number of cells are required in order to represent the deformation.

4.3.4 Motion amplitude

To determine the robustness regarding the motion amplitude between two volumes, we performed several experiments from the first US sequence detailed in table 4.4 (PHA_5). In each experiment, the inter-volume target displacement is increased by removing consecutive US volumes from original sequence. In Fig. 4.19, we show the evolution of the tracking error (y-axis) regarding the motion amplitude. As it can be seen from this figure, the tracking error remains small as long as the motion between two volumes is lower than 14 mm.

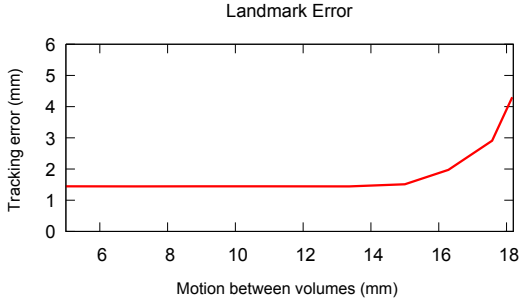


Fig. 4.19 Evolution of mean tracking error regarding motion amplitude.

4.4 Conclusion

In this chapter, we presented a real-time approach for tracking a deformable target in 3D ultrasound images. The proposed method combines intensity-based approach with physical-based model simulation. The validation was performed on phantom and real data by using the database provided by the MICCAI CLUST'14 and CLUST'15 challenges. We evaluated the robustness of the method regarding several parameters including initial segmentation error, model properties and target velocity. From these results, we observed that the proposed approach provides good performance regarding state-of-the-art methods.

However, our method remains sensitive to several issues that may have significant impact on its tracking performance. Indeed, we observed that the elasticity parameters needs to be specifically adjusted when the target undergoes high deformation (>22%). One potential solution would be to dynamically adapt these parameters from ultrasound elastography. Another issue is related to limited field of view of ultrasound images that may cause the target to be partially visible. This issue can be solved by using rectangular raw data that allows easily determining the limits of the image as demonstrated in this chapter.

While the proposed method was validated on several real sequences, it has not been evaluated regarding large shadows and intensity shift that can appear due to gain variation or US beam angle changes. In the following chapter, we propose to evaluate our approach with these artifacts. We also introduce a new similarity criterion and novel tracking strategy that allows improving the accuracy of the method.

Chapter 5

Confident-based Similarity Criteria and Strategies

In the previous chapter, we presented an approach that allows tracking deformable structures in 3D US images. While this method shows good results on phantom and real data, its performance may be perturbed by several ultrasound artifacts such as ultrasound shadows and intensity shifts. To cope with these issues, different ultrasound specific similarity criteria have been proposed in the literature. For example, Cohen and Dinstein (2002) proposed a similarity measure that assumes that the US images are degraded by Raleigh distribution noise. Baumann et al. (2012) presented a correlation-based distance measure, which is able to deal with local ultrasound intensity shift. However, these similarity criteria remain sensitive to ultrasound shadows that occlude targeted structures. Another solution consists in regularizing the displacement field from outlier detection. For example, Ni et al. (2008) provided a feature-based method that removes shadowed landmarks by detecting darkest intensity regions. Nevertheless, such approach might provide unsatisfactory results when the number of visible features is not sufficient.

In this chapter, we propose to improve our tracking method by introducing new matching criterion and novel tracking strategy based on the quality measurement of US images. From simulated data and phantom data, we demonstrate that the proposed contributions outperform classical tracking techniques since they improve robustness regarding US shadows and intensity variation. This chapter is organized as follows. In section 5.1, we describe a new similarity criterion that combines US quality measurement and Sum of Conditional Variance. In section 5.2, we describe an hybrid pairwise tracking strategy that allows improving the robustness of the tracking regarding the quality of the reference image. Finally, section 5.3 concludes this chapter.

5.1 Confident-based similarity criteria

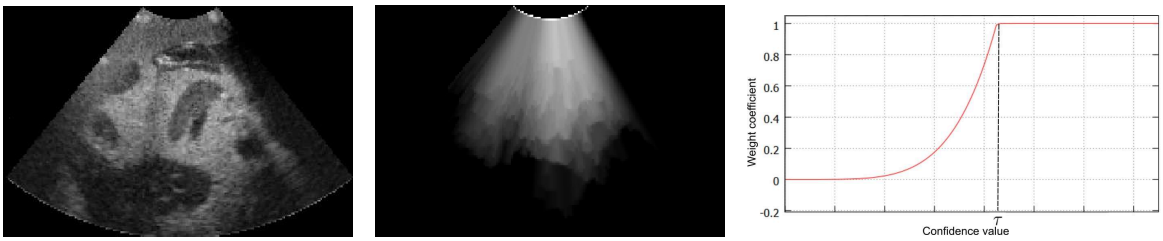
In intensity-based methods, similarity criteria play an important role since they allow measuring the alignment between images. In ultrasound applications, the SSD criterion is often used because it is simple and computationally efficient. However, it relies on the assumption that the US wave reflected by a physical point does not vary over the time. Such assumption is not always valid due to different ultrasound shortcomings. To cope with that issue, we present an ultrasound-specific matching criterion combined to shadow detection process. In section 5.1.1 and 5.1.2, we describe respectively the Weighted Sum of Squared Difference (WSSD) and the proposed Sum of Confident Conditional Variance (SCCV). In section 5.1.3, we describe experiments and results that allow comparing their performance on simulated and phantom data regarding classical similarity criteria.

5.1.1 Weighted sum of squared difference

The Weighted Sum of Squared Difference criterion consists in improving the existing SSD criterion by weighting voxels that are considered in shadowed regions. To do so, Eq. (4.7) can be modified in order to obtain a robust weighted SSD cost function as follows:

$$C(\mathbf{q}) = (\mathbf{H}_t(I_t(\mathbf{p}(t)) - I_{t_0}(\mathbf{p}(t_0))))^2 \tag{5.1}$$

where $\mathbf{H}_t$ is a $(N_v \times N_v)$ diagonal matrix that emphasize voxels that are not perturbed by US shadows. In order to define the weight matrix $\mathbf{H}_t$, several shadow detection methods have emerged (Hellier et al., 2010; Karamalis et al., 2012; Penney et al., 2004). For example, Penney et al. (2004) presented a method that remove shadow along ultrasound scanline by removing low intensity region. Another approach, proposed by Hellier et al. (2010), consists in identifying shadowed regions by detecting areas with intensity ruptures and a lower noise level. Karamalis et al. (2012) presented a technique that provides a confidence measure per voxel based on the model of ultrasound wave propagation through the tissue. We propose to use this third approach in order to detect shadows since it has been successfully implemented in recent applications such as ultrasound compounding (Berge et al., 2014) and ultrasound-based visual servoing (Chatelain et al., 2015). The approach provides, for each ultrasound image I_t, an associated confidence map U_t expressing the uncertainty measure of each voxel as it is illustrated in Fig. 5.1. From the confidence map U_t, we can directly obtain



(a) (b) (c)

Fig. 5.1 Illustration of the ultrasound image I_t (a) and its confidence image U_t (b). (c) Weighting function that provides the weighting coefficient $\mathbf{H}_t(\mathbf{p}_k(t))$ of a voxel $\mathbf{p}_k(t)$ from its confidence value $U_t(\mathbf{p}_k(t))$.

the weight matrix coefficient $\mathbf{H}_t(\mathbf{p}_k(t))$ of each voxel $\mathbf{p}_k(t)$ by using the following equation:

$$\mathbf{H}_t(\mathbf{p}_k(t)) = \begin{cases} \frac{U_t(\mathbf{p}_k(t))^\beta}{\tau^\beta}, & \text{if } 0 \leq U_t(\mathbf{p}_k(t)) < \tau \\ 1, & \text{otherwise} \end{cases} \tag{5.2}$$

where τ denotes the minimum confidence threshold that ensures the maximum weight $(\mathbf{H}_t(\mathbf{p}_k(t)) = 1)$. β is a scalar parameter that determines the smoothness of the weight matrix function. The evolution of this weighting function shape is illustrated in Fig. 5.1c.

From the Fig. 5.1c, we can therefore deduce that the matrix $\mathbf{H}_t(\mathbf{p}_k(t))$ allows discriminating unconfident voxels that have confident value lower than a certain threshold τ. This weighting matrix is not applied to the reference image since we assume that its quality is sufficient in order to obtain good tracking performance. Once the weight matrix is obtained, this criterion can be combined to the tracking method described in the previous chapter. To do so, the computation of the external displacements, expressed in Eq. (4.10), can be rewritten as follows:

$$\Delta\mathbf{q} = -\alpha\mathbf{J}^T{\mathbf{H}_t}^T\mathbf{H}_t[I_t(\mathbf{M}\mathbf{q}_t^{k-1}) - I_{t_0}(\mathbf{M}\mathbf{q}_{t_0})] \tag{5.3}$$

The performance of this criterion is evaluated in section 5.1.3.

5.1.2 Sum of confident conditional variance

The WSSD dissimilarity measure is well suited for local intensity changes but it is not robust to global intensity changes. To cope with that issue, we propose a new dissimilarity criterion based on the Sum of Conditional Variance (SCV) criterion (Pickering et al., 2009; Richa et al., 2011). Contrary to the SSD, the SCV has the advantage to be invariant to global intensity changes. Furthermore, this criterion is also computationally efficient and is therefore well

suited for real-time tracking application. The main idea of the SCV criterion consists in dynamically adapting the intensity of the reference frame over the time in order to match the intensity variation of the current frame. The SCV cost function can be expressed as follows:

$$C(\mathbf{q}) = (I_t(\mathbf{p}(t)) - \widehat{I}^t_{t_0}(\mathbf{p}(t_0)))^2 \tag{5.4}$$

where $\widehat{I}^t_{t_0}$ is a vector representing the US intensity of adapted reference frame computed at time index t. This vector is estimated by using an expectation operator $\mathcal{E}$ that takes into account the reference frame intensity I_{t_0} and the current frame intensity I_t such that:

$$\widehat{I}^t_{t_0} = \mathcal{E}(I_t|I_{t_0}) \tag{5.5}$$

Thus, the new intensity at a specific voxel position $\mathbf{p_x}(t_0)$ of the adapted reference frame is computed as follows:

$$\widehat{I}^t_{t_0}(\mathbf{p_x}(t_0)) = \sum_{i=0}^{L-1} i \, \frac{p_{I_t I_{t_0}}(i,j)}{p_{I_t}(j)} \quad \text{if } I_{t_0}(\mathbf{p_x}) = j \tag{5.6}$$

where L represents the number of grey levels of the current and reference frames. $p_{I_t I_{t_0}}$ is the joint probability density function of I_t and I_{t_0}. p_{I_t} denotes the probability density function of I_t. These functions can be computed as follows:

$$p_{I_t}(j) = \sum_{i=0}^{L-1} p_{I_t I_{t_0}}(i,j) \tag{5.7}$$

$$p_{I_t I_{t_0}}(i,j) = \sum_{k=1}^{N_v} \delta_i(I_t(\mathbf{p}_k(t)))\delta_j(I_{t_0}(\mathbf{p}_k(t_0))) \tag{5.8}$$

with $\delta_u(x)$ is an impulse function such that:

$$\delta_u(x) = \begin{cases} 1, & \text{if x=u} \\ 0, & \text{otherwise} \end{cases} \tag{5.9}$$

However, the SCV criterion, as defined by Pickering et al. (2009), is not well suited to ultrasound tracking since it is not invariant to local intensity changes that can occur due to ultrasound shadows. To tackle the previous limitation, a proposed method consists in applying the SCV criterion on sub-windows of the tracked image (Richa et al., 2014). However, these approaches may be inaccurate when the size of the window is not well

adapted to local intensity changes. To cope with that issue, we propose in this thesis a novel ultrasound-specific version of this criterion by using shadow detection.

To limit the effect of shadowed voxels, we propose to improve the previous criterion by modifying the cost function expressed in Eq. (5.4). To do so, we can refer to weighted cost function of SSD (Eq. (5.1)). The cost function can now be expressed as:

$$C(\mathbf{q}) = (\mathbf{H}_t(I_t(\mathbf{p}(t)) - \widehat{I}^t_{t_0}(\mathbf{p}(t_0))))^2 \tag{5.10}$$

The joint probability density function $p_{I_t I_{t_0}}(i,j)$ can also be perturbed by shadows since it is computed from the intensities of all voxels positions of the model. Such issue may result in mis-estimation of the new intensities of the adapted reference frame $p_{I_t I_{t_0}}(i,j)$. To solve this issue, we propose also to modify the computation of $p_{I_t I_{t_0}}(i,j)$ by emphasizing the intensity vector of voxels that are located in confident regions. To do so, the new joint probability density function can therefore be written as:

$$p_{I_t I_{t_0}}(i,j) = \sum_{k=1}^{N_v} \mathbf{H}_t(\mathbf{p}_k(t))\delta_i(I_t(\mathbf{p}_k(t)))\delta_j(I_{t_0}(\mathbf{p}_k(t_0))) \tag{5.11}$$

where $\mathbf{H}_t(\mathbf{p}_k(t))$ represents the weighted value of the voxel position $\mathbf{p}_k(t)$ that is provided by Eq. (5.2) from the confidence map. By doing so, the computation of the adapted reference frame $\widehat{I}^t_{t_0}$ is not disrupted by voxels that are located in shadowed regions. Therefore, we propose an ultrasound-specific dissimilarity measure that has the advantage to be robust to global and local intensity changes caused by presence of shadows or gain variation of the ultrasound imaging device. To link this matching criterion to the tracking method of the previous chapter, the external displacements can be computed as follows:

$$\Delta\mathbf{q} = -\alpha\mathbf{J}^T{\mathbf{H}_t}^T\mathbf{H}_t[I_t(\mathbf{M}\mathbf{q}_t^{k-1}) - \widehat{I^t_{t_0}}(\mathbf{M}\mathbf{q}_{t_0})] \tag{5.12}$$

5.1.3 Results

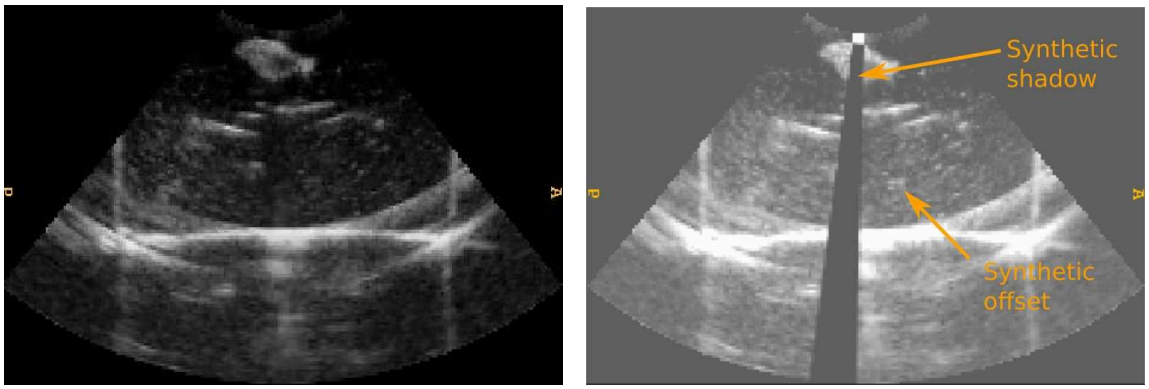
The previous experiments demonstrated the performance of the method during nominal conditions, where the region of interest is only affected by small intensity variations. However, in ultrasound-guided procedures, large shadows can appear due to the presence of bones or bad contact between skin and ultrasound probe. Furthermore, intensity variation can also be introduced due to imaging gain change of the ultrasound device. To demonstrate that our approach can cope with these issues, we evaluated our method during strong intensity changes on both simulated data and phantom data.

Results on simulated data

We first validate our method on simulated data by modifying original 3D ultrasound sequences in order to include both synthetic shadows and global intensity changes on each frame. Ultrasound shadows are simulated by replacing several ultrasound scanlines by synthetic scanlines that contain a few white voxels followed by black voxels in such a way to artificially represent a high ultrasound reflection usually observed in presence of bones. Therefore, different amounts of shadows can be obtained by changing the number of synthetic scanlines. We also simulate ultrasound gain change by adding a specific offset value on all the intensities of the ultrasound image at each frame as follows:

$$I_t(\mathbf{p}_k(t)) = I_t(\mathbf{p}_k(t)) + \kappa(t) \tag{5.13}$$

where $\kappa(t)$ is the scalar value representing the offset at time index t. In order to create a varying intensity shift, the offset value is dynamically updated over each 3D frame of an US sequence until it reaches maximum or minimum limits. An example of simulated image that includes synthetic shadow and synthetic gain change is provided in Fig. 5.2.



(a) (b)

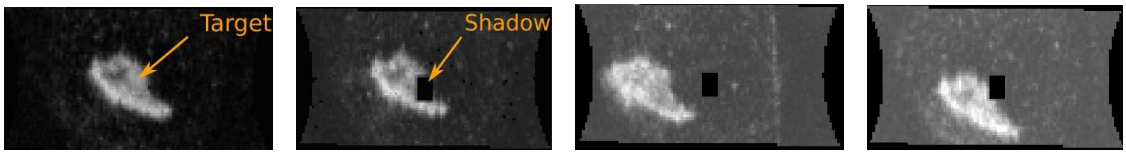
Fig. 5.2 Simulation of US imaging gain change and shadows. (a) X-Y slice of original 3D ultrasound image. (b) X-Y slice of simulated 3D ultrasound image with shadows and gain change.

For our evaluation, we generated six simulated ultrasound sequences including either rigid motion or deformation from the two original sequences. Each US sequence contains different amounts of simulated shadows and intensity variations applied to each 3D frame except the first reference image. The details of each simulated sequence are given in the table 5.1.

Sequence	Motion Type	Shadows	Offset Min/Max	Nb Frames
SIM_1	Rigid	16 lines	0/100	43
SIM_2	Rigid	40 lines	0/100	43
SIM_3	Rigid	99 lines	0/100	43
SIM_4	Non-Rigid	16 lines	0/100	100
SIM_5	Non-Rigid	44 lines	0/100	100
SIM_6	Non-Rigid	140 lines	0/100	100

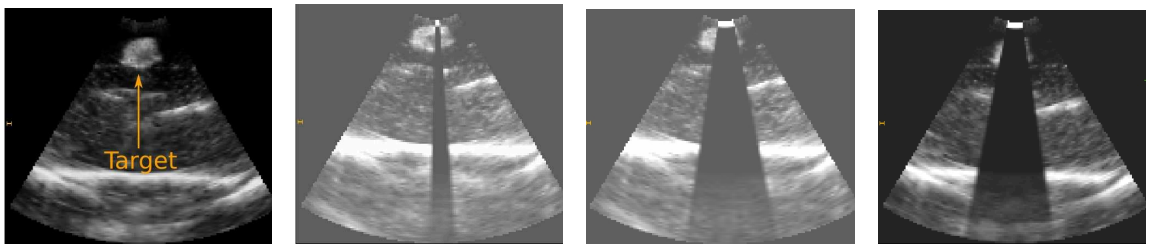
Table 5.1 Characteristics of simulated sequences. (Sequence) Name of the sequence. (Motion Type) Type of motion applied to the original sequence. (Shadows) Number of scanlines that are replaced by synthetic shadowed scanlines. (Offset Min/Max) Minimum and maximum offsets applied to ultrasound 3D frames. (Nb Frames) Number of frames in the sequence

The key frames of these simulated ultrasound sequences are illustrated in Fig. 5.3 and Fig. 5.4. Fig. 5.3 shows the first three sequences where the simulated shadow is static over the time while the target follows a rigid square motion. Consequently, the target is not affected by synthetic shadow between frame 10 and frame 26. Fig. 5.4 illustrates the last three sequences where the simulated shadow is also static but the target undergoes only a non-rigid motion. In this case, the target is affected by the shadow along each frame of the sequence.



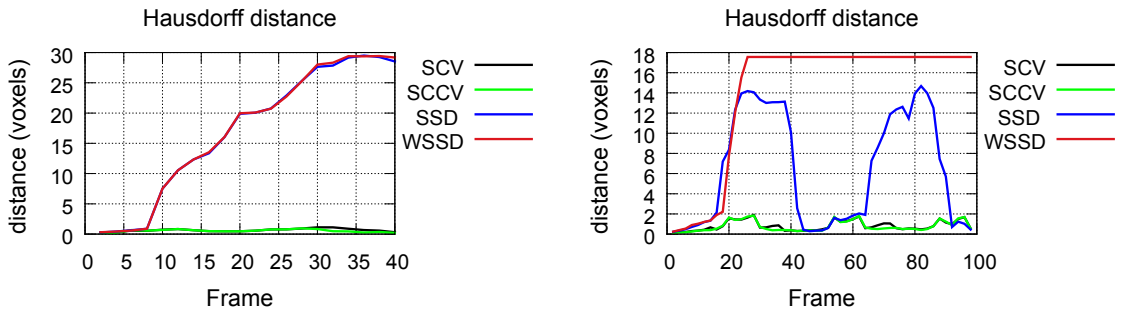
(a) (b) (c) (d)

Fig. 5.3 Illustration of sequence SIM_3 at key frames. (a) Y-Z slice of frame 0 of sequence SIM_1, SIM_2, SIM_3. (b) Y-Z slice of frame 6 of sequence SIM_3. (c) Y-Z slice of frame 15 of sequence SIM_3. (d) X-Z slice of frame 28 of sequence SIM_3.



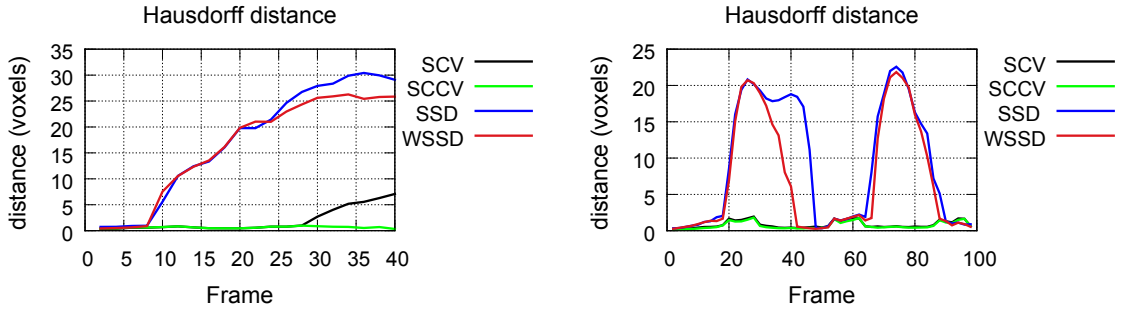
(a) (b) (c) (d)

Fig. 5.4 Illustration of sequences SIM_4, SIM_5, SIM_6 at key frames. (a) X-Z slice of frame 0 of sequence SIM_4, SIM_5, SIM_6. (b) X-Z slice of frame 25 of sequence SIM_4. (c) X-Z slice of frame 25 of sequence SIM_6. (d) X-Z slice of frame 42 of sequence SIM_6.



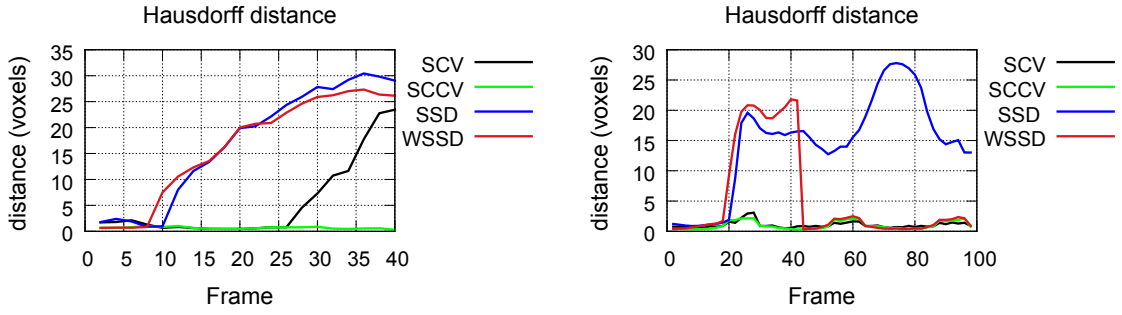
(a) Sequence SIM_1

(b) Sequence SIM_4



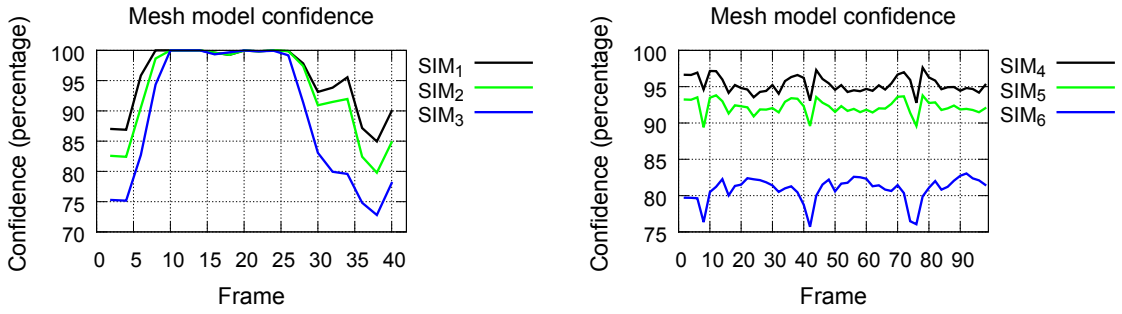
(c) Sequence SIM_2

(d) Sequence SIM_5



(e) Sequence SIM_3

(f) Sequence SIM_6



(g) Confidence evolution

(h) Confidence evolution

Fig. 5.5 Evaluation of mean tracking error of criteria on each sequence of the simulated dataset.

In order to evaluate the approach regarding the different criteria, we measured the Hausdorff distance between the model tracked in simulated sequence and the model tracked in original sequence (without shadows and gain variations). We also evaluated the percentage of confidence P_c of the mesh in order to determine how the target is affected by shadows along these simulated sequences. For that purpose, we computed the sum of confidence weights per voxel as follows:

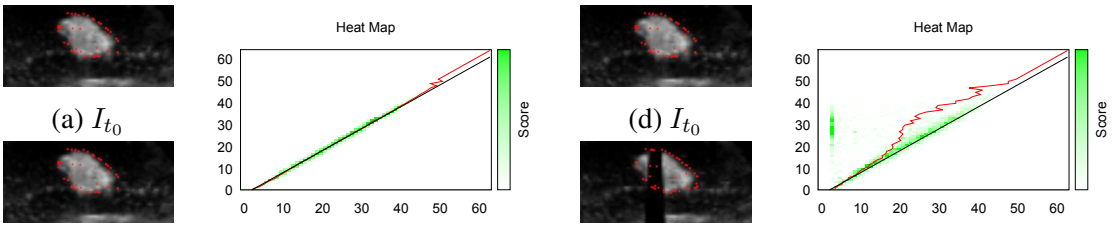
$$P_c = \frac{100}{N_v} \sum_{k=1}^{N_v} \mathbf{H}_t(\mathbf{p}_k(t)) \tag{5.14}$$

where N_v represents the total number of voxels in the mesh model. $\mathbf{H}_t$ denotes the weight matrix expressed in Eq. (5.2). If $P_c = 100\%$, then all the voxels of the model are considered confident and not affected by shadows.

In Fig. 5.5, we present the Hausdorff distance results and the model confidence along each sequence. As shown in this figure, the SSD and WSSD criteria provide inaccurate tracking for each simulated sequence. Such results were expected since these similarity measures are not well-suited during global intensity variation introduced by US gain changes. Contrary to previous criteria, SCV and SCCV present accurate results for sequences that include non-rigid motions and small confidence variation (SIM-4, SIM-5, SIM-6). Indeed, in these sequences, we can notice that the confidence evolution only slightly varies (maximum 6%). From Fig. 5.5, we can observe that our novel criterion SCCV outperforms the SCV in sequences including both rigid motions and large confidence variation (SIM-2, SIM-3). As it can be seen in Fig. 5.5, the performance of SCV criterion is related to the decrease of confidence level (after frame 26). Such result was expected because the SCV criterion is not adapted to local intensity variation introduced by shadows.

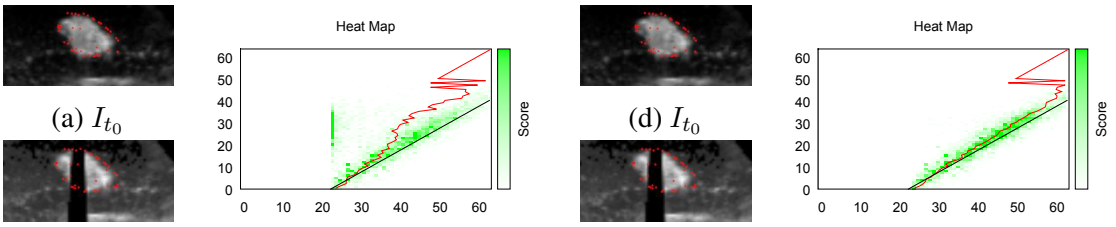
In Figs. 5.6 and 5.7, we illustrate the computation of the Joint Probability Density Function (JPDF) between the reference image and a selected current image. Let us recall that this function depends on matching criterion. Indeed, SCV criterion takes into account all the voxels for the JPDF computation, while SCCV emphasizes only confident voxels. We also propose to plot the intensity mapping function expressed in Eq. (5.6). From this latter, we observe that this function is directly obtained by averaging the JPDF weights. Furthermore, its ground truth shape can be approximated because we used simulated gain offset. Figs. 5.6c and 5.6f show respectively the JPDF computation of SCV criterion between two images. We can observe that the JPDF is very straight and narrow when strong correspondence can be established between the intensities of the initial image (Fig. 5.6a) and current image (Fig. 5.6b). The intensity mapping function is therefore also straight and fits the ground truth curve. In Fig. 5.6f, we can see that the joint probability density function is perturbed by

synthetic shadow that introduces wrong correspondences between the intensities of the initial image (Fig. 5.6d) and current image (Fig. 5.6e). Such effect can be observed in the left part of the plot. Consequently, the intensity mapping function is perturbed and does not follow the ground truth curve. In Figs. 5.7c and 5.7f, we illustrate the JPDF computation between two images obtained from SCV and SCCV criterion. The joint probability functions are more scattered since there is more noise between reference and current images. Furthermore, we can observe a horizontal shift of JPDF introduced by synthetic gain change. In Fig. 5.7c, we show the high perturbation introduced by the shadow on JPDF computation from SCV criterion. In Fig. 5.7f, this perturbation is reduced by using the SCCV criterion as it limits the effect of shadowed voxels. Consequently, the intensity mapping function from SCCV criterion has better shape and fits the ground truth curve.



(a) I_{t_0} (b) I_t (c) SCV without shadows (d) I_{t_0} (e) I_t (f) SCV with shadows

Fig. 5.6 Comparison of the estimation of joint probability density function $p_{I_t I_{t_0}}$ during tracking task. The x-axis and y-axis represent respectively the intensity level of the target at the current frame and at the original reference frame. The green cloud represents the joint probability density function. The black and red curves represent respectively the intensity conversion function estimated from ground truth and from joint probability density function. (a) SCV criterion applied to original frame 2 without shadows (b) SCV criterion applied to simulated frame 2 with synthetic shadows.



(a) I_{t_0} (b) I_t (c) SCV: shadows/gain change (d) I_{t_0} (e) I_t (f) SCCV: shadows/gain change

Fig. 5.7 Comparison of the estimation of joint probability density function $p_{I_t I_{t_0}}$ during tracking task. The x-axis and y-axis represent respectively the intensity level of the target at the current frame and at the original reference frame. The green cloud represents the joint probability density function. The black and red curves represent respectively the intensity conversion function estimated from ground truth and from joint probability density function.(c) SCV criterion applied to simulated frame 74 with shadows and imaging gain change (d) SCCV criterion applied to simulated frame 74 with shadows and gain change.

From the previous simulation results, we can conclude that in the presence of imaging gain variation, the SSD and WSSD criteria produce inaccurate results while the SCV and SCCV provide robustness by adapting the reference frame over the time. However, in the presence of shadows, we can also observe that the SCV criterion provides good results as long as the shadow and motion remain small. We can also notice that the new criterion tackles the previous issues by limiting the effect of the shadowed voxels.

Results on phantom data

We also evaluated our approach regarding the different similarity criteria on two phantoms. For that purpose, we acquired ultrasound sequences that contain real shadows and/or gain variation. The details of each sequence are provided in table 5.2. To evaluate the accuracy of the different criteria, we annotated the position of a specific target landmark for 10 frames of each ultrasound sequence. Thus, we can compare the ground truth landmark positions and the warped point positions estimated from our tracking approach. The overall tracking results are shown in table 5.3. In the following experiments, specific landmarks are difficult to identify due to the lack of texture within targets. To cope with that issue, we choose respectively either the far left point or the far right point of the targets as landmarks.

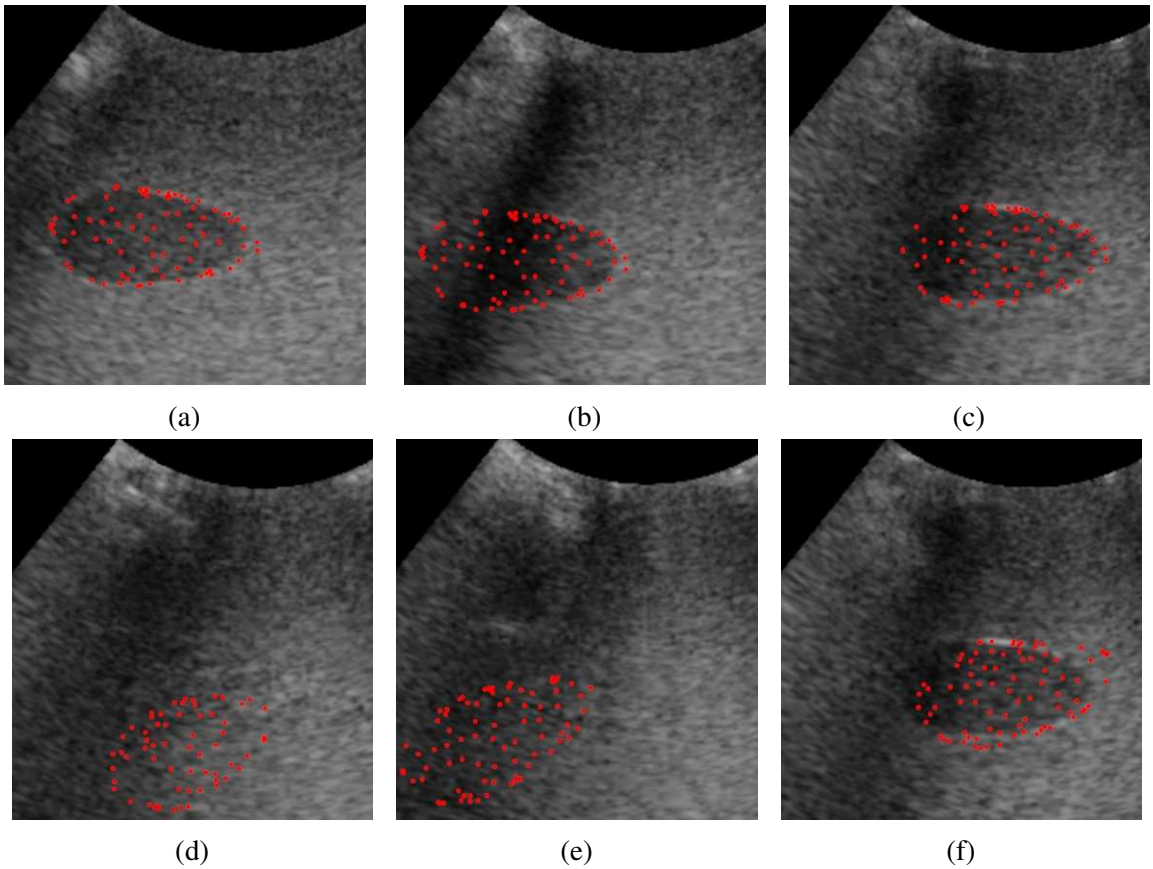
Sequence	Motion Type	Shadows	Gain change	Nb Frames
PHA_1	Translation	Yes	No	100
PHA_2	Rotation	Yes	No	100
PHA_3	None	No	Yes	100
PHA_4	Translation	Yes	Yes	100

Table 5.2 Characteristics of phantom sequences. (Sequence) Name of the sequence. (Motion Type) Type of motion applied to the original sequence. (Nb Frames) Number of frames in the sequence.

Sequence	SSD	WSSD	SCV	SCCV
PHA_1	10.6 ± 11.7	5.8 ± 6.8	22.9 ± 31.1	2.48 ± 2.18
PHA_2	-	5.36 ± 6.01	–	2.0 ± 1.7
PHA_3	42 ± 46	–	1.7 ± 1.2	–
PHA_4	–	–	31 ± 41	2.4 ± 2.2

Table 5.3 Accuracy evaluation of similarity criteria regarding each US sequence. The results are expressed in millimeters and represent respectively the mean tracking error $\pm$ the standard deviation (Sequence) Name of the sequence.

The first sequence PHA_1 allows comparing all the similarity criteria regarding the presence of large shadows and translation motion. Fig. 5.8 illustrates the tracking tasks on several frames and show that the SSD and SCV criteria diverge due to the amount of shadows that occludes the target. We can also notice that the WSSD criterion shows better results but the model does not fit exactly the target surface over the time. Contrary to the other criteria, the SCCV criterion provides robust tracking along the ultrasound sequence. It can be also seen in table 5.3, where we can observe that the SCCV provides smallest mean tracking error. In Fig. 5.9, we illustrate the evolution of the model confidence where we see that the target is strongly occluded by the shadow since its confidence decreases down to 30% around frame 50.



(a) (b) (c)

(d) (e) (f)

Fig. 5.8 Example of tracking tasks during the PHA_1 sequence with shadows and translation motion. Red points represent the intersections between the mesh model and the Y-Z slice of 3D image. (a-b-c) SCCV tracking at frame 2 (a), at frame 52 (b), at frame 99 (c) of sequence PHA_1. (d) SSD tracking at frame 99 , (e) SCV tracking at frame 99, (f) WSSD tracking at frame 99. The different images correspond to the Y-Z slices that passes through the barycenter of the tracked mesh over the time. Therefore, the target can not be seen in Figs. (d) and (e) since SSD and SCV criteria diverged.

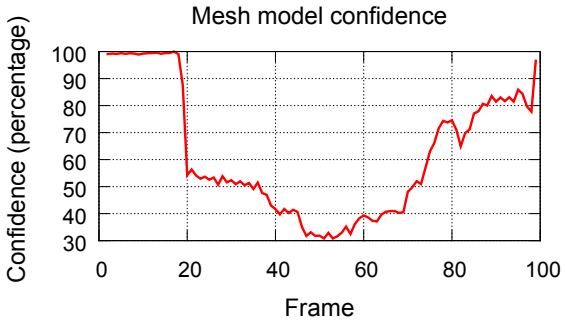


Fig. 5.9 Evolution of the mesh model confidence during the sequence PHA_1. We can observe that the confidence of the mesh model decreases down to 30% at frame 50.

The second experiment consists in evaluating our approach on an ultrasound sequence that contains shadows and rotation motion. In this experiment, we only compare the performance of the WSSD and SCCV criteria since we demonstrate that they provide better results regarding the presence of shadows. Fig. 5.10 shows the tracking tasks on several frames. We can notice that the SCCV criterion provides more accurate tracking than WSSD criterion since its model fits the target surface over the time. This can be also seen in table 5.3 that shows that the SCCV criterion achieves a smaller mean tracking error than the WSSD criterion.

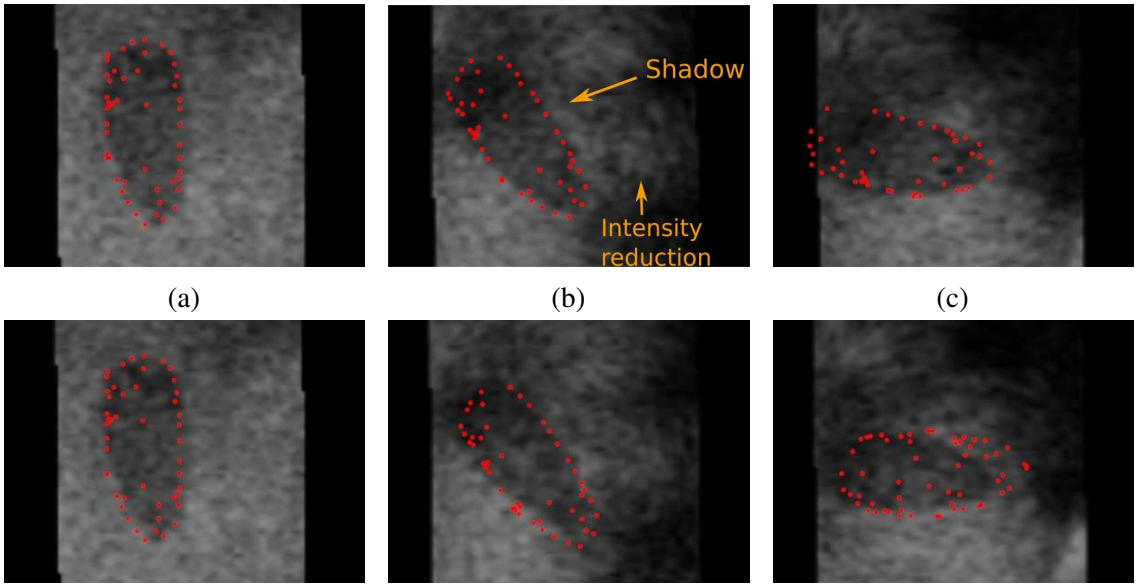


Fig. 5.10 Example of tracking tasks during the PHA_2 sequence with shadows and rotation motion. Red points represent the intersections between the mesh model and the Y-Z slice of 3D image. (a-b-c) SCCV tracking (a) at frame 2, (b) at frame 122, (c) at frame 145 of sequence PHA_2. (d-e-f) WSSD tracking (d) at frame 2, (e) at frame 122, (f) at frame 145 of sequence PHA_2.

In the sequence PHA_3, we compare the performance of our approach only regarding global intensity changes. For this purpose, we acquired an ultrasound sequence that does not contain any motion and we modify the ultrasound gain during the acquisition. It is worth mentioning that the target is not displaced since no motion is applied in this sequence. We only compare the SCV and SSD criteria since the target undergoes only global intensity changes. The tracking tasks are illustrated on some frames in Fig. 5.11 where we can notice that the SSD tracking task diverges since the model is displaced over the time and it does not fit the target surface. We can also observe that the SCV criterion achieves considerable better results than SSD from the table 5.3.

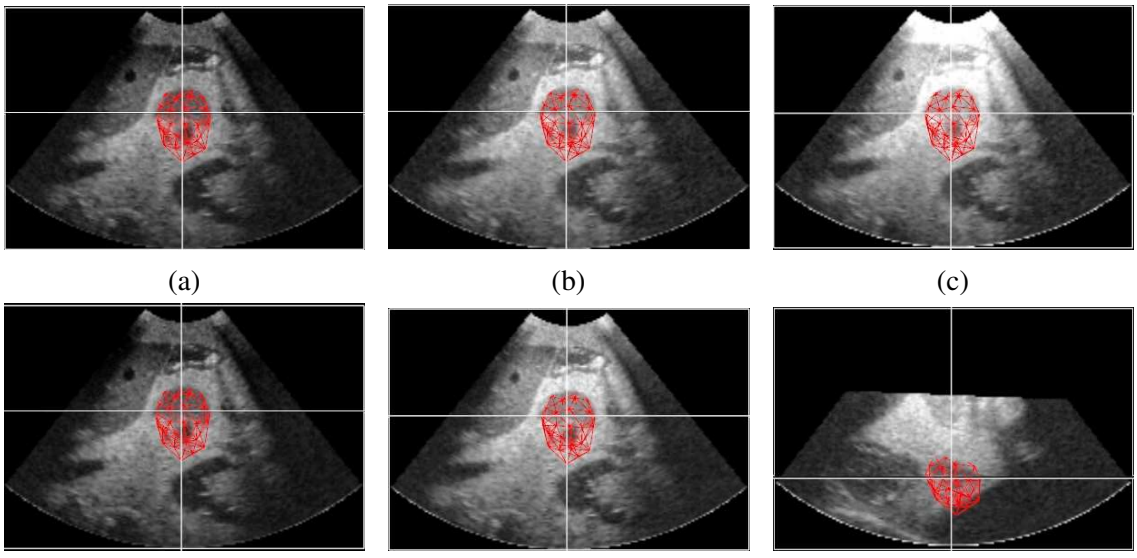
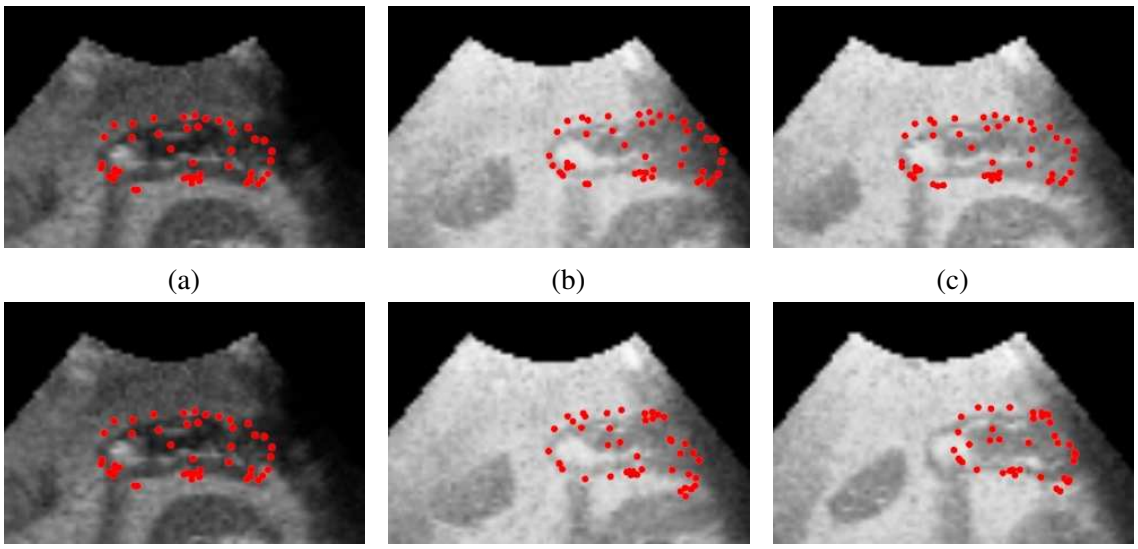


Fig. 5.11 Example of tracking tasks during the PHA_3 sequence with gain variation. Red model represents the 3D mesh model. (a-b-c) SCV tracking (a) at frame 2, (b) at frame 8, (c) at frame 40 of sequence PHA_3. (d-e-f) SSD tracking (d) at frame 2, (e) at frame 8, (f) at frame 40 of sequence PHA_3. The different images correspond to the X-Y slices that passes through the barycenter of the tracked mesh over the time. Therefore, the target can not be seen in Fig. (f) since SSD criterion diverged.

The last experiment consists in testing our approach on an ultrasound sequence containing both shadows and imaging gain variation. Furthermore, we also introduced a translation motion by manually moving the probe. We only compare the performance of the SCV and SCCV criteria since we demonstrated in the previous experiment that the SSD and WSSD criteria are not robust to global intensity changes. The tracking tasks are illustrated on several frames in Fig. 5.12. From this figure, we can observe that the SCV criterion diverges when the target is displaced in shadowed region after frame 50. We can also notice that our new criterion provides robust results during local and global intensity changes with a final mean tracking error of 2.4 mm from table 5.3. Fig. 5.13 illustrates the evolution of the mesh model confidence during the tracking task. We can observe that the SCV tracking diverges when the target confidence decreases down to 80% after frame 50.



(a) (b) (c)

(d) (e) (f)

Fig. 5.12 Example of tracking tasks during the PHA_4 sequence with local and global intensity variations. Red points represent the intersections between the mesh model and the image slice. (a-b-c) SCCV tracking (a) at frame 2, (b) at frame 51, (c) at frame 71 of sequence PHA_4. (d-e-f) SCV tracking (d) at frame 2, (e) at frame 51, (f) at frame 71 of sequence PHA_4.

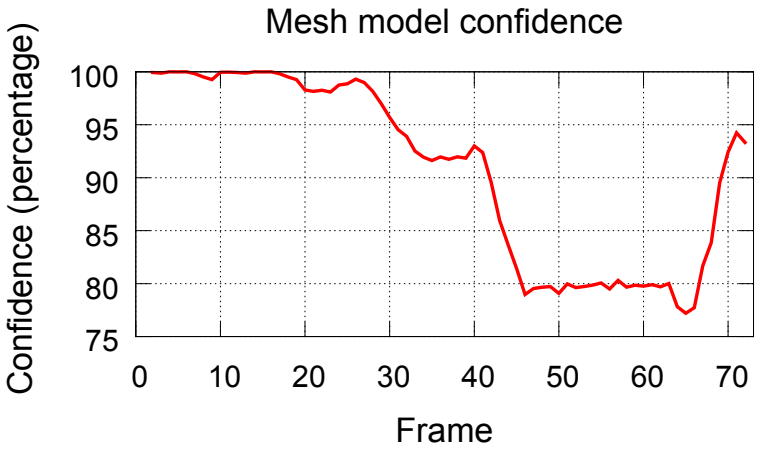


Fig. 5.13 Evolution of the mesh model confidence during the sequence PHA_4.

From the previous experiments, we can deduce that the SSD and WSSD criteria are sensitive to global intensity shifts. To cope with that issue, SCV and SCCV criteria are more robust to ultrasound gain changes by adapting the reference target intensity. However, the SCV is not well adapted to local intensity variation that can occur due to the presence of ultrasound shadows. From the sequences PHA_1 and PHA_2, we also notice that the WSSD can not provide accurate results against shadows. This is caused by the intensity reduction effect in the vicinity of the shadow as illustrated in Fig 5.10(b). Finally, we can notice that our new criterion SCCV is robust to local and global intensity changes even if large shadows significantly occludes the target (70% of the target in experiment PHA_1 in Fig. 5.9). The performance of the proposed criterion open novel perspectives since it allows tracking structures that are affected by strong ultrasound artifacts. In the following section, we propose a new tracking strategy based on SCCV criterion that allows maximizing the quality of reference image.

5.2 Confident-based tracking strategy

In previous section, we proposed a new criterion that is robust to local and global intensity variation introduced by ultrasound artifacts. However, the main issue of SCCV criterion is related to the assumption that the quality of the reference image is high enough. Such assumption is not always valid and may have significant impact on the tracking performance. To cope with that issue, we propose a tracking strategy based on previous criterion. In the literature, the tracking strategies can be distinguished into two main categories, i.e.

pairwise and groupwise strategies. In the following sections, we give a description of each type of strategy and we detail their advantages and shortcomings. We also propose a new hybrid tracking strategy based on the quality measurement of ultrasound images. Finally, we demonstrate that our method outperforms other strategies through simulated and phantom data.

5.2.1 Pairwise strategy

In ultrasound tracking applications, pairwise approach is the most used since it relies on classical registration scheme between current and reference images. These methods can be subdivided into iterative and non-iterative pairwise strategies depending on how the reference image is selected. Iterative pairwise strategies aim at finding the optimal transformation by registering consecutive frames together. In ultrasound application, examples of these methods have been proposed by Shekhar and Zagrodsky (2002) and Heyde et al. (2012). Therefore, for each new acquired image, iterative pairwise strategy defines the spatial transformation T by minimizing the following cost function:

$$\arg\min_{T} S(I_{t-1}, T \circ I_t) + R(T) \tag{5.15}$$

where I_t and I_{t-1} represent consecutives images acquired at time t and $t-1$. However, registering consecutive pairs of images can lead to error accumulation and drift over the time. To cope with that issue, several authors propose non-iterative strategies that consist in registering all the images to a selected reference image. Therefore, non-iterative strategies aim at finding the spatial transformation T that minimizes the following cost function:

$$\arg\min_{T} S(I_r, T \circ I_t) + R(T) \tag{5.16}$$

where I_r and I_t represent respectively the selected reference image and the current image. The main shortcoming of these strategies is related to the critical choice of the reference image I_r. In most of tracking applications, the initial US image of the sequence is selected as the reference template. However, this image quality should be high enough in order to ensure good tracking performance. Furthermore, the appearance of both reference and current images should be similar in order to obtain an unique optimal transformation. To cope with that issue, several methods propose to periodically update the reference image when the error between reference and current images is higher than a threshold.

5.2.2 Groupwise strategy

A second type of methods, named groupwise strategies, consists in finding the optimal displacement of moving structure from the simultaneous alignment of multiple images. Therefore, these approaches take into account all image information in order to lead to more robust registration results. Contrary to pairwise methods, these strategies are invariant to the bias error introduced by the selection of a reference image. In ultrasound applications, examples of groupwise methods have been proposed by Metz et al. (2011) and Vijayan et al. (2013). Metz et al. (2011) proposed a method that consists in registering a current frame and an implicit frame obtained from previous images. The estimation of the target displacement is obtained by minimizing the following cost function:

$$\arg\min_{T} S(\widehat{I_m}, T \circ I_t) + R(T) \tag{5.17}$$

where $\widehat{I_m}$ is an implicit frame computed from previous images of the sequence. It allows eliminating the need to choose a reference image. Vijayan et al. (2013) propose to compute the implicit frame by averaging the intensity of the previous image over the time as follows:

$$\widehat{I_m} = \sum_{t_k=t_0}^{t-1} I_{t_k}(\mathbf{p}(t_k)) \tag{5.18}$$

where I_{t_k} represents the intensity vector of the image at time index k. $\mathbf{p}(t_k)$ denotes a vector that defines all the voxels positions of the target at time index k. From the previous equation, we can see that this approach makes the assumption that intensity values at corresponding spatial locations over time are constant. Compared to pairwise strategies, the main limitation of these strategies is related to its computational complexity. Indeed, the execution time of the proposed approach increases with the sequence length. Therefore, this type of strategy is not well suited for real-time tracking of long sequence, as it requires a huge amount of memory and processing power. In the table 5.4, we summarize the advantages and shortcomings of each tracking strategy.

Type of Strategy	**Quality Invariant**	**Drift Invariant**	**Comp. time**
Iterative Pairwise	✓	✗	✓
Non-iterative Pairwise	✗	✓	✓
Groupwise	✓	✓	✗

Table 5.4 Characteristics of existing tracking strategies. (Comp. time) Computation time of the method

5.2.3 Hybrid strategy

In this section, we introduce a new tracking strategy that allows tackling the issues of the previous approaches. The proposed method is based on non-iterative pairwise strategy that registers a current image with a selected reference image. However, as stated previously, the main limitation of this strategy is related to the reference image selection since this image can be affected by large shadows. Such issue may introduce bias error during tracking task because it reduces the number of confident voxels. To cope with that issue, we propose to locally adapt the reference image over the time. Contrary to iterative strategies, the proposed method aims at replacing only unconfident areas of the selected template in order to limit error accumulation. By doing so, the proposed approach ensures the maximum quality of the reference image. This strategy computes the spatial transformation by minimizing the following cost function:

$$\arg\min_{T} S(I_d, T \circ I_t) + R(T) \tag{5.19}$$

where I_d is a dynamic image that depends on previous images of the sequence. Our strategy aims at obtaining the dynamical reference image I_d from quality measurement of previous US images as follows:

$$I_d(\mathbf{p}_{t_0}) = \begin{cases} I_{t_0}(\mathbf{p}_{t_0}), \text{ if } t = t_0 \\ I_t(\mathbf{p}_t), \text{ if } U_t(\mathbf{p}_t) > U_d(\mathbf{p}_{t_0}) \end{cases} \tag{5.20}$$

where I_t and I_{t_0} represent respectively the current and the initial frames. U_t denotes the current confidence image, while U_d is the dynamic confidence image representing the quality of the image I_d. It is obtained from previous confidence images and can be expressed as follows:

$$U_d(\mathbf{p}_{t_0}) = \begin{cases} U_{t_0}(\mathbf{p}_{t_0}), \text{ if } t = t_0 \\ U_t(\mathbf{p}_t), \text{ if } U_t(\mathbf{p}_t) > U_d(\mathbf{p}_{t_0}) \end{cases} \tag{5.21}$$

The proposed method has the advantage to be faster than groupwise strategies as it relies only on the alignment of two images. Furthermore, such approach prevents the error accumulation as it locally replaces structures until the quality of the reference image is sufficient. Finally, this strategy avoids the bias error introduced by the selection of reference image.

5.2.4 Results

In this section, our novel tracking strategy is evaluated from two different types of experiments. First, we compare the tracking approaches through simulated data that consist of original US sequences that are degraded by simulated shadows. The second type of experiments allows evaluating the tracking strategies on phantom data. It it worth mentioning that only pairwise strategies are compared since groupwise strategy is not well-suited for real-time applications.

Results on simulated data

The simulation experiments are performed on three original US sequences that are perturbed by different amounts of simulated shadows. Contrary to previous simulation experiments, the initial image of each sequence is also affected by ultrasound shadows. In Table 5.5, we detail some characteristics of each sequence, e.g. the motion type and the amount of simulated shadow.

Sequence	Motion Type	Shadows	Nb Frames
SIM_1	Non-Rigid	140 lines	43
SIM_2	Rigid	16 lines	43
SIM_3	Rigid	40 lines	43

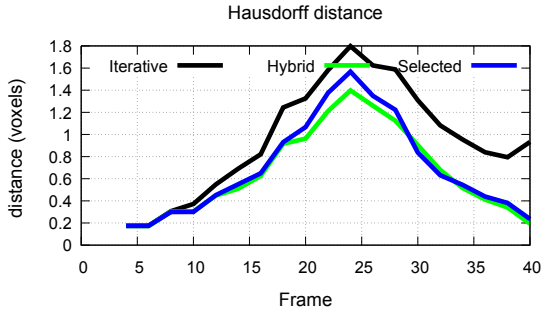
Table 5.5 Details of simulated sequences. (Sequence) Name of the sequence. (Motion Type) Type of motion applied to the original sequence. (Shadows) Number of scanlines that are replaced by synthetic shadowed scanlines. (Nb Frames) Number of frame in the sequence

To evaluate the performance of the strategies, we measure their accuracy and their bias error along each sequence. To do so, we compare the Hausdorff distance between the models tracked in perturbed sequence and in original sequence. The ground truth is provided by results obtained from original sequence since this latter is not perturbed by shadows. We also evaluate the bias error by measuring the intensity difference between the reference image and original reference image. This error can therefore be expressed as follows:

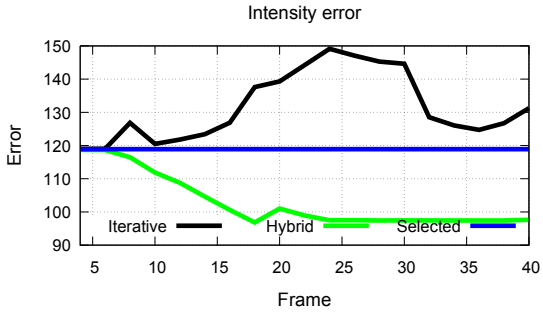
$$E_{bias} = ||^{o}I_r(\mathbf{p}(t_0)) - {}^{p}I_r(\mathbf{p}(t_0))||^2 \tag{5.22}$$

where ${}^{o}I_r$ denotes the initial image in the original sequence. ${}^{p}I_r$ denotes the reference image in the simulated sequence. Depending on the chosen tracking strategy, the intensity vector ${}^{p}I_r$ of the reference image may vary. For example, non-iterative pairwise tracking strategy provides constant reference image such that ${}^{p}I_r = {}^{p}I_{t_0}$. Iterative strategy aims at changing

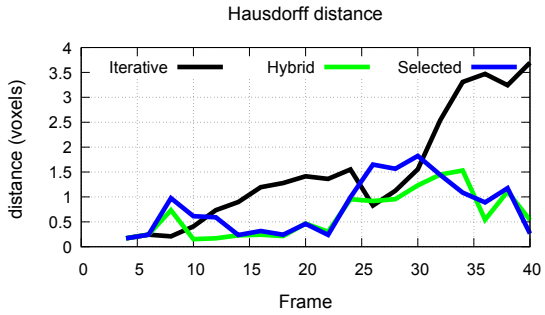
the reference image when a new image is acquired such that ${}^{p}I_r = {}^{p}I_{t-1}$. The hybrid strategy allows locally adpating the reference image until its quality is sufficient such that ${}^{p}I_r = {}^{p}I_d$.



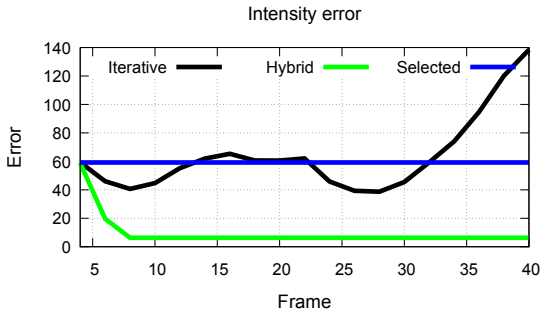
(a) Sequence SIM_1 : Distance Error



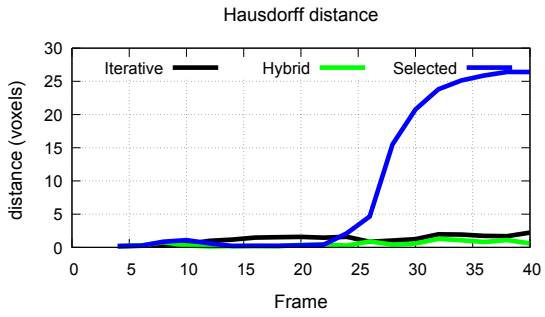
(b) Sequence SIM_1 : Intensity Bias error



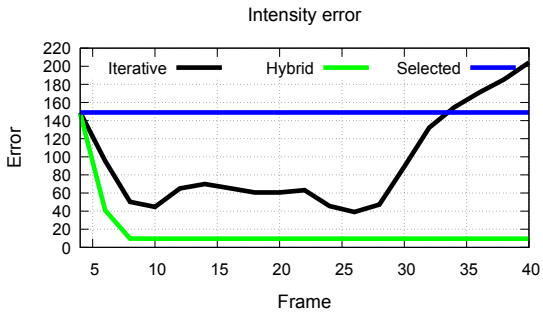
(c) Sequence SIM_2 : Distance Error



(d) Sequence SIM_2 : Intensity Bias Error



(e) Sequence SIM_3 : Distance Error



(f) Sequence SIM_3 : Intensity Bias Error

Fig. 5.14 Comparison of pairwise tracking strategies on each sequence of the simulated dataset. (Iterative) Iterative strategy, (Hybrid) Proposed Hybrid strategy, (Selected) Non-Iterative strategy with selected reference image.

The results of these experiments are presented in Fig. 5.14. From that figure, we can notice that for each simulation, non-iterative and iterative strategies show respectively constant and unstable evolution of bias error over each frame. This difference may be explained by the adaptation of reference image from iterative strategy. However, such technique may lead to error accumulation as it can be seen in Figs. 5.14d and 5.14f after frame 25. Contrary to other strategies, we can observe that our method allows progressively reducing the bias error of the reference image since it locally adapts the reference image until its quality is sufficient. In terms of tracking accuracy, we can observe that hybrid and non-iterative strategies outperforms iterative strategy for sequence SIM_1 and SIM_2. Indeed, this latter is sensitive to error accumulation and therefore may lead to tracking drift over the time. We can also notice that our strategy provides better tracking results for simulation SIM_3. The bad tracking performance of non-iterative strategy is related to the poor quality of reference image. Indeed, this strategy may provide incorrect results because it does not allow updating the reference image.

Results on phantom data

The second type of experiment allows comparing the tracking strategies through 3D ultrasound images of synthetic phantom. To do so, we acquired a sequence whose quality varies over the time by introducing US shadows. These latter are obtained from bad contact between phantom and US probe. In this experiment, the tracking accuracy of each strategy is obtained by measuring the tracking error with respect to an annotated landmark over each image. The ground truth is provided from manual annotations of one expert over each frame.

The tracking task is illustrated in Fig. 5.16 from which we can observe that the initial image of the sequence is highly perturbed by shadows. In Fig. 5.15, we compare the tracking accuracy of each strategy over the time. We can observe that non-iterative pairwise strategy provides high tracking error along the sequence with a mean tracking error of 5.65 mm. Such result may be explained by the poor quality of reference image that reduces considerably the number of confident voxels. We can also notice that iterative strategy provides better results (MTE= 3.78 mm) since it allows updating the reference image along the sequence. However, as illustrated in Fig. 5.15, we can see that the tracking error of the strategy increases at the end of the sequence due to error accumulation. Therefore, we can see that the proposed strategy outperforms existing pairwise strategies by providing stable and low tracking error along the sequence (MTE =1.37 mm). However, we can also notice that high error can be obtained at specific frames, e.g. frame 30, frame 83, and frame 127. This may be induced by several factors like manual annotation inaccuracy, or high inter-frame motion. Nevertheless,

as it can be seen from Fig. 5.15, the method remains accurate along the sequence by limiting error accumulation.

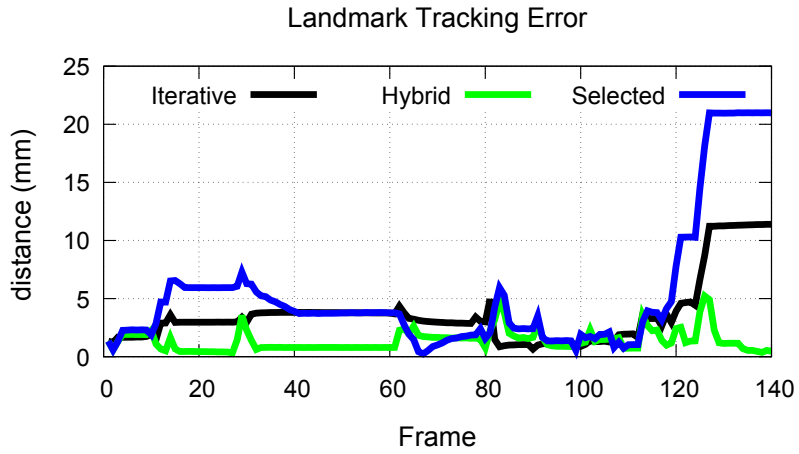


Fig. 5.15 Tracking error results obtained from Phantom experiment. Comparison of different pairwise tracking strategies. (Iterative) Iterative strategy, (Hybrid) Proposed Hybrid strategy, (Selected) Non-Iterative strategy with selected reference image.

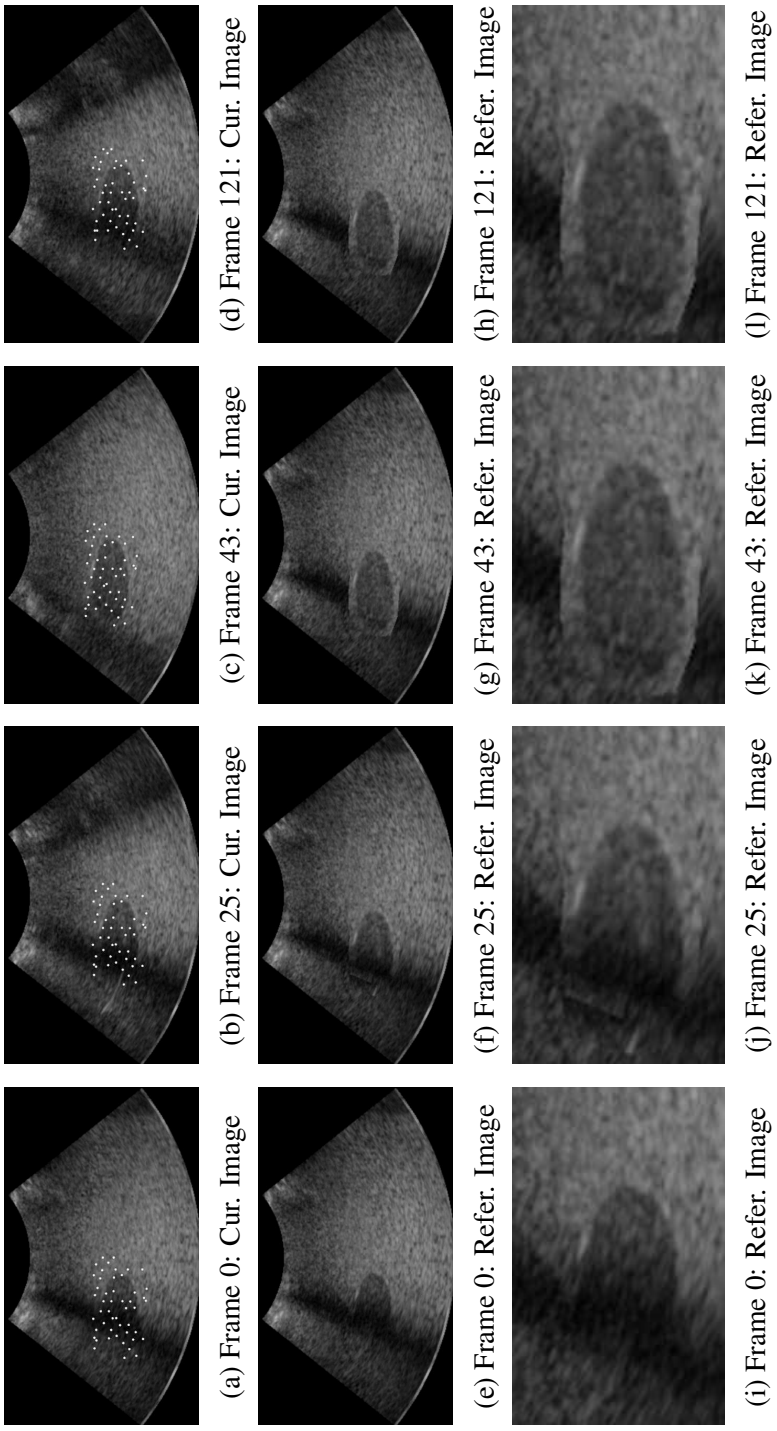


Fig. 5.16 Illustration of tracking results with hybrid strategy. (Left column) Current images during tracking task. White points represent the intersections between the mesh model and the Y-Z slice of 3D image. (Middle column) Dynamic Reference image I_d during tracking task. (Right column) Zoom on reference model during tracking task.

5.3 Conclusion

In this chapter, we presented a novel matching criterion that combines Sum of Conditional Variance and quality measurement of US images. Such criterion has the advantage to be robust to local and global intensity variations that can occur due to large shadows and gain variation. We demonstrated the proposed criterion outperforms classical matching criteria from simulation and phantom experiments. Therefore, the proposed approach allows opening novel perspectives in ultrasound-guided therapies when the visualization of targeted structures is affected by ultrasound artifacts. However, this method relies on the assumption that quality of the reference image is sufficient in order to ensure correct tracking results.

To cope with that issue, we introduced a new hybrid pairwise tracking strategy based on the SCCV criterion. The proposed approach allows locally adapting the reference image until its maximum quality is reached. Such approach has the advantage to be invariant to bias error introduced by the selection of reference image. Furthermore, it is more robust to tracking drift that may be induced by error accumulation. From simulation and phantom experiments, we show that our hybrid strategy is more accurate than existing pairwise strategies. Finally, our method is less computationally demanding than groupwise strategies because it only relies on the alignment of two images.

In the following chapter, we propose to evaluate these contributions on real-data by using an ultrasound sequence of liver tissue acquired on a volunteer. In addition, we propose a new visualization system that combines pre-operative and 3D ultrasound images. Such system relies on accurate tracking of structures in 3D US images by combining SCCV criterion and mechanical-based regularization.

Chapter 6

Application to Multi-modal Visualization

In previous chapters, we introduced several contributions that allow tracking deformable structures in ultrasound images even if they are affected by several ultrasound shortcomings including speckle noise, gain variation, and shadows. As it can be seen from the results, the proposed tracking approach achieves better performance than classical tracking techniques. The robustness and accuracy of the proposed contributions allow opening novel perspectives in multi-modal registration, visualization and medical robotics. In this chapter, we proposed a clinical application that allows improving surgeon visualization by combining pre-operative image and ultrasound images. To do so, we present a new visualization method that replaces the shadowed areas of US images by visible structures from pre-operative images. These areas represent structures that can not be seen in US images due to the strong reflections of the ultrasound signal. They are generally caused by solid structures such as bones, but can also occur due to bad contact between skin and ultrasound probe. To evaluate the proposed application, the visualization system is tested in clinical practice. In section 6.1, we describe the acquisition of 3D US sequence and MRI volume. The whole method including calibration and registration steps is presented in section 6.2. In section 6.3, we evaluate the approach through an experiment performed in real clinical scenario.

6.1 Materials

In this experiment, we acquired a 3D US sequence and MRI data of a healthy volunteer at the University Hospital of Rennes. The MR volume is obtained from a 3T Siemens Verio MRI research system and its associated dimensions are 640 x 520 x 64 voxels with a related spacing of 0.59 x 0.59 x 3.0 mm. The ultrasound volumes are obtained from an Ultrasonix station thanks to a motorized probe (4DC7-3/40). This probe incorporates a curvilinear transducer array mounted on a motor that sweeps in forward and backward direction in

order to acquire adjacent 2D images. Then, each 3D ultrasound volume is constructed by interpolating the data between the acquired 2D frames. Each US volume is generated from 11 adjacent 2D images with a motor angular step of 2.19°. The dimensions of the US volumes are respectively of $781 \times 518 \times 221$ voxels with an associated voxel spacing of $0.30 \times 0.30 \times 0.30$ mm. Thus, the US field of view is equal to $24.09°$ and its associated volume rate is 0.5 vol/s.

6.2 Method

In this section, we propose a method that allows improving US visualization by replacing shadowed regions of US images by structures visible in pre-operative image. Such approach has the advantage to not disturb the surgeon visualization since only regions that do not represent any clinical structures are removed from the US images. The proposed approach can be decomposed into three steps:

- **Multi-modal Registration:** The first step aims at registering the initial US volume and pre-operative image.
- **Ultrasound Tracking:** Once the initial registration is performed, a region of interest is segmented in pre-operative image and tracked in the remaining of the sequence.
- **Pre-operative/Ultrasound visualization:** The last step aims at providing an output image that replaces unconfident regions of the current US image by structures visible in pre-operative image.

6.2.1 Multi-modal registration

This step consists in aligning MRI volume and the initial 3D US images. This registration problem can be represented as follows:

$$E(T) = \arg\min_{T} E(I_{us}, T \circ I_{po}) \tag{6.1}$$

where I_{us} and I_{po} represent respectively the initial ultrasound image and the pre-operative image. T denotes the transformation applied to the source image. In this approach, T is considered as a rigid transformation characterized by translation and rotation parameters. E is a cost function that quantifies the level of alignment between the images. To perform registration between two images, several solutions are available including manual, semi-automatic, and automatic multi-modal registration methods. In our approach, the registration is obtained

from external registration by using localization system. Then, a manual registration step is performed in order to refine the final transformation. The calibration of the different imaging systems is required in order to have the different imaging modalities represented in the same space coordinate system. In the following sections, we describe respectively the experimental setup, and registration and calibration steps.

Experimental setup

In this section, the experimental setup that allows externally registering MR and Ultrasound images is described. In Fig. 6.1, we illustrate this setup where $\mathcal{F}_{ius}$ and $\mathcal{F}_{imr}$ represent respectively the coordinate systems associated to the US and MRI volumes. To fuse the different imaging modalities, the classical solution consists in initializing the registration from external localization system. Such system allows retrieving the 3D position and orientation of passive markers with respect to the coordinate frame of localization system $\mathcal{F}_o$. As illustrated in Fig. 6.1, a passive marker is attached to the ultrasound probe in order to detect the position of the US volume. However, any passive marker can be linked to the MRI system since it is not compatible with electromagnetic perturbations introduced by localization system. To cope with that issue, several mri-compatible fiducials are stuck to the volunteer skin as illustrated in Figs. 6.1 and 6.2. After the acquisition, the positions of these fiducials are recorded from passive marker attached to a touch probe. The coordinate systems of the touch probe and US probe are denoted as $\mathcal{F}_t$ and $\mathcal{F}_p$. From this experimental setup, an initial transformation can be retrieved from the external registration.

External registration

The external registration is based on an external localization system. To do so, several devices can be used including electromagnetic, optical, or computer vision systems. In this application, we chose to use optical system since it is robust and invariant to electromagnetic perturbations. This system is equipped with a position sensor along with an infrared light emitter. The passive markers are detected since they are composed of infrared-reflective spheres. The objective of external registration is to find the rigid transformation ${}^{imr}\mathcal{H}_{ius}$ that links the MRI coordinate frame $\mathcal{F}_{imr}$ to the US coordinate frame $\mathcal{F}_{ius}$. This rigid transformation ${}^{imr}\mathcal{H}_{ius}$ can be computed as follows:

$$ {}^{imr}\mathcal{H}_{ius} = {}^{o}\mathcal{H}_{imr}^{-1}\,{}^{o}\mathcal{H}_{ius} \tag{6.2} $$

where ${}^{o}\mathcal{H}_{ius}$ denotes the rigid transformation that relates the US image coordinate frame to the optical camera coordinate frame. ${}^{o}\mathcal{H}_{imr}$ represents the rigid transformation that links

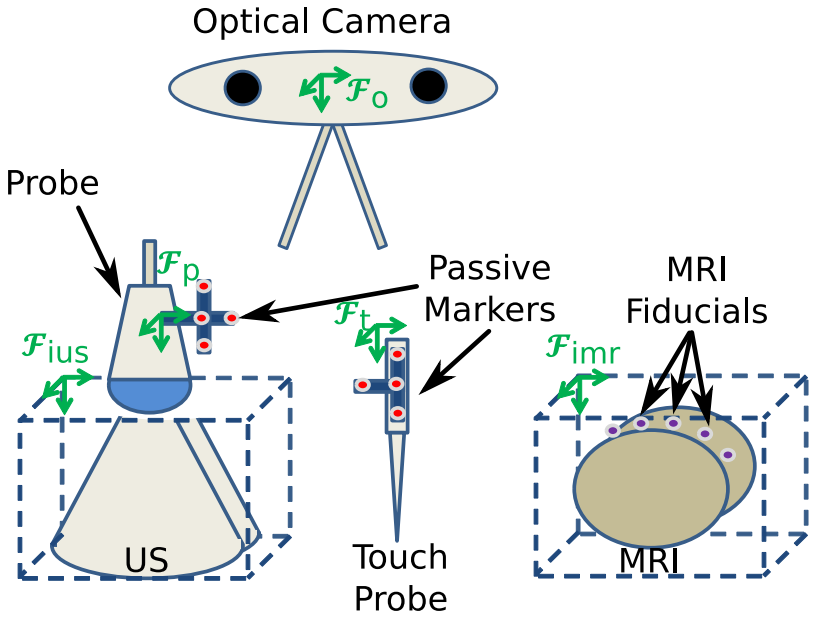


Fig. 6.1 Illustration of the acquisition protocol. $\mathcal{F}_o$ represents the frame associated to the optical camera frame. $\mathcal{F}_{ius}$ and $\mathcal{F}_{imr}$ represent the coordinate systems associated to the US images and MRI images. $\mathcal{F}_p$ and $\mathcal{F}_t$ are the frames associated to the touch probe and US probe.

the MR image coordinate system to the optical camera coordinate system. As it can be seen from Fig. 6.1, only the transformations denoted as ${}^{o}\mathcal{H}_p$ and ${}^{o}\mathcal{H}_t$, relating both the US probe frame and touch probe frame to the optical camera frame, can be retrieved from the localization system. Therefore, in the following, we describe the calibration steps that allows obtaining the transformations ${}^{o}\mathcal{H}_{ius}$ and ${}^{o}\mathcal{H}_{imr}$. Once the imaging systems are calibrated, we can compute the transformation ${}^{imr}\mathcal{H}_{ius}$ by rewriting the Eq. (6.2) as follows:

$$ {}^{imr}\mathcal{H}_{ius} = {}^{o}\mathcal{H}_{imr}^{-1}\,{}^{o}\mathcal{H}_{p}\,{}^{p}\mathcal{H}_{ius} \tag{6.3} $$

MRI calibration

The calibration of MRI system aims at providing the rigid transformation ${}^{o}\mathcal{H}_{imr}$ that links the optical camera frame $\mathcal{F}_o$ with the MRI frame $\mathcal{F}_{imr}$. In order to find this transformation, we propose to relate the positions of several mri-compatible fiducials between these two coordinate systems. To do so, several water-based fiducials are attached to the volunteer skin as illustrated in Fig. 6.2. Therefore, such fiducials can be segmented by using an intensity-threshold since they appear bright within MRI volume. In the optical camera coordinate system, the positions of these fiducials are recorded by using the touch probe. Once the coordinates of these fiducials are recorded in both frames, we can compute the rigid

transformation ${}^{o}\mathcal{H}_{imr}$ that minimizes the following least square cost function:

$$e(\mathbf{R},\mathbf{t}) = \frac{1}{N_f}\sum_{i=0}^{N_f}(\mathbf{p}^f_{imr}(i) - (\mathbf{R}\mathbf{p}^f_o(i) + \mathbf{t}))^2 \tag{6.4}$$

where $\mathbf{p}^f_{imr}(i)$ and $\mathbf{p}^f_o(i)$ represent respectively the homogeneous coordinates of the i-th fiducial in MRI and optical camera frames. e denotes the least square cost function. N_f is the number of fiducials points. The translation vector $\mathbf{t}$ and the rotation matrix $\mathbf{R}$ represent the parameters of the cost function that can be found by using the singular value decomposition algorithm. The transformation ${}^{o}\mathcal{H}_{imr}$ can thus be obtained by combining the previous parameters as follows:

$${}^{o}\mathcal{H}_{imr} = \begin{pmatrix} \mathbf{R}_{3\times3} & \mathbf{t}_{3\times1} \\ 0 \quad 0 \quad 0 & 1 \end{pmatrix} \tag{6.5}$$

Once the MRI system is calibrated, it remains to define the rigid transformation ${}^{o}\mathcal{H}_{ius}$ from the calibration of the US imaging system in order to compute the external rigid registration.

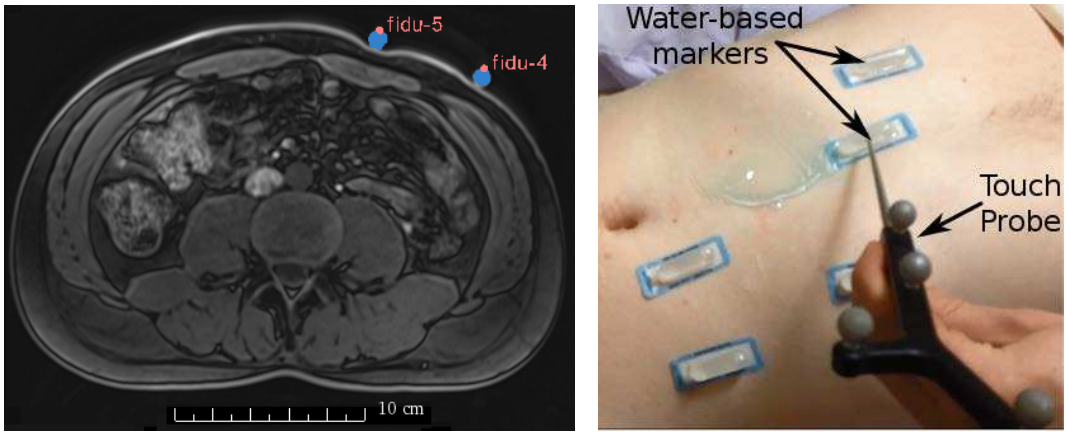


Fig. 6.2 Illustration of MRI calibration steps. (Left) Axial slice of MR volume of the healthy volunteer. The blue circles represent the segmentation of fiducials. (Right) Image representing the acquisition of the fiducials positions by the touch probe.

US Calibration

To perform external registration, the calibration of US imaging system is required since it allows providing the transformation ${}^{p}\mathcal{H}_{ius}$ required in Eq. (6.3). In our application, the calibration of the 3D US probe is performed by using an algorithm based on N-wires phantom

from the Plus library proposed by Lasso et al. (2014). This technique relies on the detection of fiducials points in US images representing the intersections between several wires and 2D US image plane. This set of wires is attached to a calibration phantom whose geometry is known. Then, the calibration method consists in determining the transformation that relates the coordinates of fiducials points in both US image frame and probe frame. As this method is adapted to 2D image, we developed a framework that allows calibrating 3D ultrasound volumes from the extraction of the 2D image corresponding to the central observation plane of the 3D probe. The calibration algorithm for 3D ultrasound volume can be decomposed into three steps:

- **Data Collection and Volume Rendering:** This step consists in collecting each 2D frame of a volume as well as the associated transformation matrices that relate the optical camera frame to the frames of passive markers attached to both the ultrasound probe and the calibration phantom. Once the volume is complete, the volume is rendered by using bi-linear interpolation between 2D acquired data. The volume generation is illustrated in Fig. 6.3a.

- **Central Frame Extraction:** As the US volume geometry is known, the calibration of one 2D frame is sufficient in order to calibrate the whole volume. To do so, we retrieve only the central image of the rendered volume as illustrated in Fig. 6.3b.

- **Points Detection:** Once the central image is extracted, the last step consists in detecting and segmenting the bright image points representing the intersections between middle US frame and the phantom wires. An example of US image during calibration is shown in Fig. 6.3c. Then, an optimization algorithm is performed in order to find the transformation that relates the probe frame to the 2D central frame.

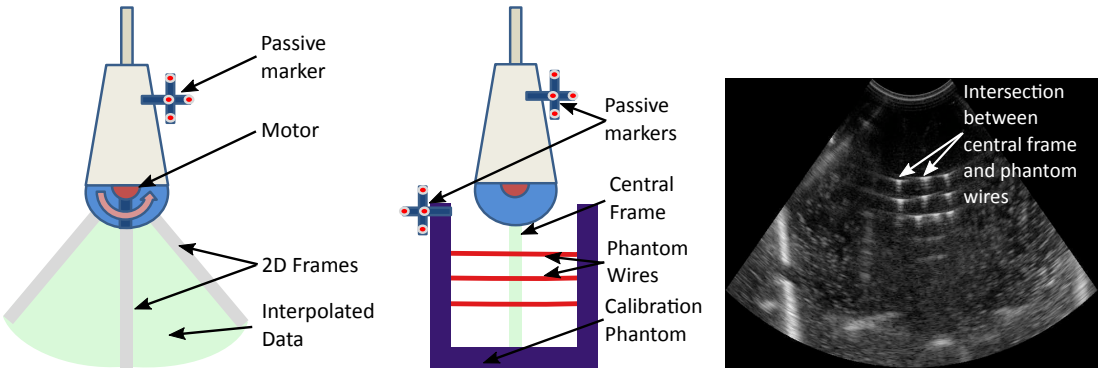


Fig. 6.3 Illustration of the 3D ultrasound volume calibration. (a) Illustration of Data Collection and Volume Rendering. (b) Central Frame Extraction and phantom calibration. (c) Example of US image during calibration

Once the imaging systems are calibrated, an initial registration between MR and initial US image can be computed thanks to the Eq. (6.3). However, the external registration may suffer from non-negligible inaccuracies due to the physiological motions of the patient. To cope with that issue, other registration techniques can be combined in order to refine registration between MR and US images.

Manual registration

In this work, we chose to refine the final registration thanks to manual registration from the annotations of corresponding features in both images. Therefore, the final transformation T can be expressed as follows:

$$T = {}^{imr}\mathcal{H}_{ius}\mathcal{H}_{man} \tag{6.6}$$

where $\mathcal{H}_{man}$ represents the manual transformation that allows refining the registration between MR image and initial US image. To obtain this transformation, we annotate landmarks between initial US volume and warped MRI images as illustrated in Fig. 6.4. Once these annotations are performed, the manual transformation can be obtained by decreasing the following cost function:

$$e(\mathbf{R}^m, \mathbf{t}^m) = \frac{1}{N_f}\sum_{i=0}^{N_f}(\mathbf{p}_{wmr}(i) - (\mathbf{R}^m\mathbf{p}_{ius}(i) + \mathbf{t}^m))^2 \tag{6.7}$$

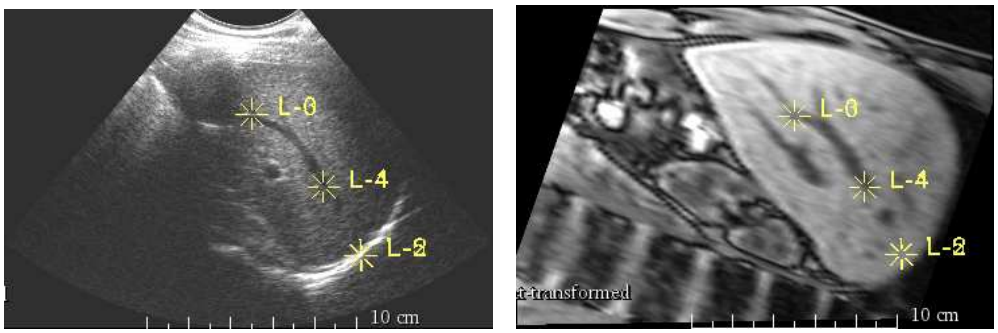
where $\mathbf{p}_{wmr}$ represents the landmarks annotated in warped MRI image as illustrated in Fig. 6.4. $\mathbf{p}_{ius}$ denotes the landmarks annotated in initial US image. $\mathbf{R}^m$ and $\mathbf{t}^m$ respectively represent the rotation matrix and translation vector computed from minimization of the previous cost function. The manual transformation $\mathcal{H}_{man}$ can thus be obtained by combining the previous parameters as follows:

$$\mathcal{H}_{man} = \begin{pmatrix} \mathbf{R}^m_{3\times3} & \mathbf{t}^m_{3\times1} \\ 0 \quad 0 \quad 0 & 1 \end{pmatrix} \tag{6.8}$$

The main issue associated to manual registration is the time comsumption due to the user interaction. To tackle this limitation, automatic registration methods can also be used as described by Coupé et al. (2012).

6.2.2 Ultrasound tracking

Once the registration between initial US volume and MRI volume is performed, a region of interest is segmented in pre-operative image and a corresponding mesh model is generated.



(a) (b)

Fig. 6.4 Illustration of manual annotations. (a) Manual annotations in initial US image. (b) Manual annotations in warped MRI image obtained from transformation ${}^{imr}\mathcal{H}_{ius}$.

This model contains a set of cells linked by a set of nodes and allows tracking the region of interest in the remaining images of the ultrasound sequence. To do so, we used the contributions described in previous chapters that consists in the combination of robust similarity metric (SCCV) and mechanical-based regularization. Thus, the node positions $\mathbf{q}(t)$ at time t of the model can be obtained by using the following equation:

$$\mathbf{q}(t) = \mathbf{q}(t-1) + h_i \Delta\mathbf{q} + \Delta\mathbf{d} \tag{6.9}$$

where $\mathbf{q}(t-1)$ denotes the previous positions at time $t-1$. $\Delta\mathbf{d}$ is the internal node displacements obtained from the integration of mechanical model forces. $\Delta\mathbf{q}$ represents the external node displacements obtained from the optimization of the similarity criterion. The term h_i represents a gain that amplifies the effect of external displacements estimation.

6.2.3 Pre-operative/Ultrasound visualization

The last step aims at providing an output image that replace unconfident regions of the current US image by structures visible in pre-operative image. The output image can be computed as follows:

$$I_v(\mathbf{p}(t)) = \begin{cases} I_t(\mathbf{p}(t)), & \text{if } U_t(\mathbf{p}(t)) > \Theta_c \\ I_{po}(\mathbf{p}(t_0)), & \text{otherwise} \end{cases} \tag{6.10}$$

where I_t and I_{po} represent respectively the current ultrasound image and the pre-operative image. U_t and I_v denote respectively the ultrasound confidence image and the output image. Θ_c is the confidence threshold that determines if the voxel is in shadow region, and $\mathbf{p}(t_i)$ represents the voxel positions at time index t_i.

6.3 Results

Multi-Modal registration

The registration and calibration errors are summarized in table 6.1. Thanks to the external localization system, we can perform external registration between the MRI and initial US volumes from the Eq. (6.3). The initial registration error is approximately 19.2 mm and is thus too large in order to be used directly in the visualization application. This large error can be caused by several factors including respiratory motion and ultrasound probe pressure. To tackle this issue, we propose to compensate this error by adding a manual transformation. The final error after manual registration is below 2 mm and allows obtaining good initialization of the visualization system.

Protocol	**Mean Error**
Optical Loc. Precision	0.30
Ultrasound calibration	1.22
MRI calibration	4.91
External registration	19.23
Final registration (+ manual)	1.80

Table 6.1 Registration and calibration errors expressed in millimeters.

Ultrasound tracking

Once the initial US volume and MR image are registered, we propose to track the region of interest along the ultrasound sequence by using the proposed tracking approach described in previous chapters. In order to provide ground truth, the position of a specific landmark representing hepatic vein bifurcation was annotated over each ultrasound image of the sequence. The accuracy of different similarity criteria is evaluated on the same tracking task. The results of the mean tracking error are summarized in table 6.2 and show that SCCV criterion has the advantage to propose greater accuracy performance. Other criteria provide lower performance as the ultrasound sequence is affected by ultrasound shadows and beam angle changes inducing local intensity variation.

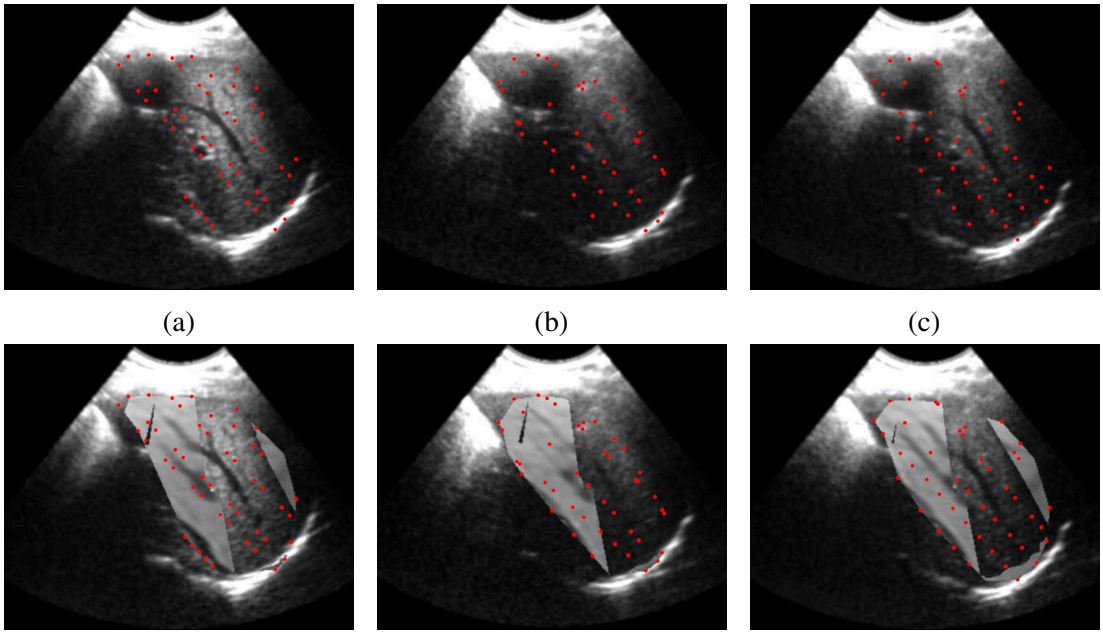
Confident-based visualization

The previous tracking and registration results allow setting up the visualization system as their error is under 2.4 mm. Some of the results are illustrated in Fig. 6.5. As it can be seen

Similarity criterion	Mean Tracking Error
Sum of Squared Differences	7.40
Weighted Sum of Squared Differences	3.82
Sum of Confident Conditional Variance	2.32

Table 6.2 Tracking errors expressed in millimeters

from the results, the sequence is perturbed by different amount of shadows between frames 0 and 21. The proposed application allows improving visualization by replacing occluded structures of ultrasound imaging by warped pre-operative MRI imaging.



(a) (b) (c)

(d) (e) (f)

Fig. 6.5 Example of the tracking and visualization tasks. The top images represent X-Y slice of US volumes (a) at time index 0, (b) at time index 14, (c) at time index 21. The bottom images represent X-Y slice of output volumes (d) at time index 0, (d) at time index 14, (e) at time index 21. Red points represent the intersection between 3D mesh model and image plane.

6.4 Conclusion

Ultrasound imaging is often used in order to guide percutaneous interventions because it has the advantage to propose real-time visualization of soft tissue. However, during these procedures, surgeon visualization is often perturbed by US shadows that occlude some clinical structures of interest due to strong reflection of US waves. Consequently, the accuracy and the intervention time of these procedures may be impacted. To cope with that issue, we propose an application that allows improving the surgeon visualization by replacing shadowed structures of ultrasound images by pre-operative data. We performed an experiment from which we show that our new application may be implemented in a real clinical scenario by using a multi-modal registration framework. From the previous results, we can see that the proposed tracking approach may open novel perspectives in intra-operative interventions guided by ultrasound images.

Chapter 7

Conclusion

7.1 Initial objectives

With the development of medical technologies, the use of minimally-invasive therapies considerably increased with respect to classical surgery. These interventions allow delivering treatment with greater surgical precision and therefore reduce pain and recovery time of the patient. In this thesis, the focus is set on minimally-invasive therapies for liver tumor ablation. These treatments consist in eliminating hepatic tumors by inserting a needle that delivers heat energy or alcohol around malignant regions. While these therapies have progressed dramatically in the past years, several issues still have an impact on their effectiveness. A first type of problem is related to the required accuracy of these interventions since they can only treat small regions around the needle tip. Such problem may force the surgeon to readjust the needle position because targeted structures may be displaced by physiological motions and medical tools manipulation. To cope with that issue, ultrasound imaging guidance is often used since it provides real-time visual feedback of the tissues. However, this modality suffers from several shortcomings such as speckle noise, shadows, and limited field of view. These latter may complicate the visualization of targeted structures or needle. To overcome this limitation, different tracking approaches have been proposed since they allow estimating the position of targeted structures over the time. Despite the number of approaches presented in the literature, it is still difficult to assess the performance of these methods as they are evaluated on different databases. Furthermore, these approaches are often validated on nominal conditions from US sequences that do not contain any shadows and intensity variation. However, in minimally-invasive procedures, these artifacts are generally present and significantly affect the US images. Therefore, such problem can be an obstacle to the implementation in clinical practice.

7.2 Achieved work

This thesis provides several contributions that allow obtaining accurate and robust estimation of target displacements in US sequence. Furthermore, in this work, we also introduced an application that fuses a sequence of US images with a pre-operative volume. This section gives a brief description of our work.

Physically-based tracking approach

We first proposed an approach that combines intensity-based method and mechanical-based regularization. In this method, the target displacements are computed by iteratively summing both external and internal displacements over the time. The internal displacements are obtained from simulation of mass-spring-damper system associated to the target model. The external displacements are computed from dense motion estimation. We showed that our approach has the advantage to follow accurately motions even if the targeted structure is affected by high speckle noise. We also provided an adaptation of this method that allows using ultrasound raw data acquired along scanlines. The proposed adaption avoids reconstructing each 3D US volume of the sequence and allows tracking partially visible targets. From evaluation on phantom and real data, we showed that the proposed method outperforms state-of the-art approaches. To complete this study, we also evaluate the robustness of the method with respect to mechanical parameters, initial segmentation error, and target displacement.

US-specific similarity criterion

Among our contributions, we also proposed an ultrasound-specific matching criterion that combines quality measurement of US image and Sum of Conditional Variance (SCV) criterion proposed by Pickering et al. (2009). For this purpose, we used an expectation operator that computes the joint probability density function between reference and current images. This function is emphasized only on confident regions that are not affected by shadows. The main benefits of our similarity measure include its low computation time and its robustness regarding local and global intensity variation. From simulated and phantom data, we demonstrated its excellent performance on challenging ultrasound sequences that are affected by gain variation and large shadows. In these experiments, we also showed that our matching criterion is more accurate than classical similarity criteria.

Hybrid tracking strategy

The last contribution of this thesis is related to a novel tracking strategy. In the literature, classical tracking strategies aim at estimating target displacements from comparison between two or more images. For example, pairwise strategies perform registration between successive pair of images. However, these methods are sensitive to the quality of the selected reference image. To cope with that issue, groupwise strategies have been introduced since they align the whole sequence at the expense of the computational cost. In this thesis, we proposed a novel strategy that aims at registering current image and dynamical reference image. The method allows updating the intensity of reference image until its quality is sufficient. Our approach has the advantage to be fast and robust to tracking drift induced by reference image selection. From simulated and phantom data, we show that our approach is more robust than existing pairwise strategies. Furthermore, the proposed strategy is not computationally demanding contrary to groupwise strategies since it only relies on the comparison of two images.

Application to multi-modal visualization

Thanks to the contributions presented previously, we provided a tracking approach that obtains accurate and robust estimation of target displacements in sequence of ultrasound volumes. The results demonstrate that our approach can handle different US artifacts such as speckle decorrelation and shadows. To evaluate this method in a real clinical scenario, we proposed an application that allows improving surgeon visualization. This application aims at providing complete visualization of soft tissues by fusing US images and pre-operative image. Such method relies on the combination between multi-modal registration framework and the tracking contributions described previously. To evaluate this application, an experiment fusing MRI and US imaging was performed at the University Hospital of Rennes.

7.3 Future work

In this thesis, we presented a tracking approach that combines several contributions in order to obtain robust estimation of target position in a sequence of 3D ultrasound images. However, they are limitations that still prevent the usage of this method in a real clinical practice. Several perspectives, that may help solving these issues, are also discussed in this section.

Patient-specific model

In this thesis, we presented a tracking approach that relies on mechanical-based regularization obtained from the simulation of mass-spring-damper system. Despite its excellent performance on different databases, some inaccuracies can be observed when the target deformation is too high. We showed that this issue may be related to the setting of mechanical parameters that are not well adjusted with respect to the organ motion. To cope with this problem, a potential solution could be to dynamically tune the parameters of mechanical model from elastography data. Such approach can provide better tracking results since it would adapt the regularization term with respect to soft tissues of the patient. Furthermore, it is worth mentioning that elastography data can be extracted from ultrasound images without any additional cost. Another perspective would be to replace the mass-spring-damper system by a more realistic model. For example, Haouchine et al. (2013) proposed an augmented reality application that considers a finite element method based on a co-rotational formulation. Such model is more adapted to the liver movements since it allows large displacements while relying on a linear expression of the stress-strain relationship.

Automatic segmentation

As stated previously, the proposed tracking method relies on manual segmentation of the target within the first volume of the US sequence. Nevertheless, such task is not adapted to real clinical practice since it requires a certain amount of time due to manual interaction. Therefore, future work could consist in tracking objects that are automatically segmented from 3D US images. For that purpose, we can use methods proposed by Chang et al. (2005) or Barbosa et al. (2014). Such methods may offer average segmentation errors up to 2.29 and 2.26 mm. However, pre-operative images can also be used as they provide better results thanks to their greater imaging quality and their larger field-of-view. For example, Heimann et al. (2009) proposed a method that provides a Root Mean Square Symmetric Surface Distance of 1.4 mm.

Evaluation during treatment

In this work, we showed that the proposed contributions provide sufficient accuracy and robustness in order to estimate the position of targeted structures. These results have been obtained from simulated, phantom and real data acquired from different volunteers. However, a future work could consist in evaluating our approach on US sequences acquired during per-cutaneous minimally-invasive treatments. Indeed, during these procedures, several effects may perturb the tracking performance. A first issue is related to the needle insertion that

introduces organ deformation until a crack is formed in the tissue. Furthermore, high local intensity variation can be observed around the needle position. To take into account this artefact, an interesting approach would be to detect the needle within ultrasound sequences. To do so, Uhercik et al. (2009) proposed an interesting approach based on random sample consensus algorithm. The tracking performance can also be affected by the thermal or chemical destruction that generate coagulated lesion around the targeted tissue. Such effect introduces hyper-echogenic area during the ablative process. To tackle this problem, Seo et al. (2011) proposed an approach where coagulated lesion is considered as landmark during the tracking task. Such landmark has interesting properties since it is rigid and has a speckle pattern that is distinct from the background. Finally, the tissue necrosis may also produce gas during the treatment. Such effect generates US shadows that lead to inaccurate assessment of treatment response and calculation of ablation zone dimensions. To cope with that issue, a potential solution would be to evaluate the proposed SCCV criterion during the treatment.

Sophisticated optimization strategy

In this thesis, we observed that the proposed method may provide inaccurate results when the target displacement is high between consecutive volumes. Such problem may complicate the implementation in a real clinical scenario. It is worth recalling that the target displacement between consecutive volumes is related to the acquisition frame rate of the used probe. Therefore, a potential solution would be to use recent ultrasound systems providing high acquisition rate. However, this solution requires reducing the computation time of the proposed approach so that it can remain below the acquisition time of our ultrasound system. To do so, several potential solutions may be evaluated. A first solution consists in reducing the size of ultrasound volumes by using down-sampling operators. However, such solution may have an impact on the tracking accuracy. A more sophisticated approach would be to use other optimization approaches such as inverse compositional operators as proposed by Delabarre and Marchand (2014). Such method allows reducing the computational cost since the Jacobian matrix is kept constant throughout the tracking task. This strategy could improve tracking accuracy if it is combined to more evolved optimization strategies such as Levenberg-Marquardt or efficient second order minimization schemes.

Automatic registration

In this work, we proposed a novel multi-modal application that combines ultrasound images and magnetic resonance imaging. While the obtained results are promising, the proposed application suffers from different shortcomings that may complicate its usage in clinical

practice. A first problem is related to initial registration computed from external localization system. Such system provided a coarse alignment between images by detecting the position of passive markers on the US probe and on the patient's skin. However, the usage of such method is difficult to implement in clinical scenario since it requires a particular attention on the markers placement in order to ensure their visibility. A potential solution consists in initializing the registration from markerless registration as proposed by Gilles et al. (2016). This method detects the position of different patients from time of flight cameras. Once the initialization is performed, the registration is refined from manual registration obtained from annotations of landmarks of one expert. However, this step may also complicate the procedure since it requires time and efforts due to the manual interaction. To cope with that issue, one method can consist in obtaining the initial registration from automatic method as proposed by Coupé et al. (2012).

References

(2007). *Numerical recipes: the art of scientific computing*. Cambridge University Press, Cambridge, UK ; New York, 3rd ed edition.

Angelova, D. and Mihaylova, L. (2010). Contour segmentation in 2d ultrasound medical images with particle filtering. *Machine Vision and Applications*, 22(3):551–561.

Baker, S. and Matthews, I. (2004). Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255.

Banerjee, J., Klink, C., Peters, E. D., Niessen, W. J., Moelker, A., and van Walsum, T. (2015). Fast and robust 3d ultrasound registration – Block and game theoretic matching. *Medical Image Analysis*, 20(1):173–183.

Barbosa, D., Friboulet, D., D'hooge, J., and Bernard, O. (2014). Fast tracking of the left ventricle using global anatomical affine optical flow and local recursive block matching. *Proc. of the MICCAI Challenge on Endocardial Three-dimensional Ultrasound Segmentation-CETUS*, pages 17–24.

Basarab, A., Liebgott, H., Morestin, F., Lyshchik, A., Higashi, T., Asato, R., and Delachartre, P. (2008). A method for vector displacement estimation with ultrasound imaging and its application for thyroid nodular disease. *Medical Image Analysis*, 12(3):259–274.

Baumann, M., Mozer, P., Daanen, V., and Troccaz, J. (2012). Prostate biopsy tracking with deformation estimation. *Medical image analysis*, 16(3):562–576.

Bell, M. A. L., Byram, B. C., Harris, E. J., Evans, P. M., and Bamber, J. C. (2012). In vivo liver tracking with a high volume rate 4d ultrasound scanner and a 2d matrix array probe. *Physics in Medicine and Biology*, 57(5):1359–1374.

Berge, C. S. z., Kapoor, A., and Navab, N. (2014). Orientation-Driven Ultrasound Compounding Using Uncertainty Information. In *Proc. of International Conference on Information Processing in Computer-Assisted Interventions*, volume 8498, pages 236–245, Cham. Springer International Publishing.

Bookstein, F. L. (1989). Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585.

Chagnon, S., QANADLI, S., and LACOMBE, P. (2001). Destruction percutanée des tumeurs du foie par radiofréquence. *Gastroentérologie clinique et biologique*, 25:B85–B99.

Chang, R.-F., Wu, W.-J., Moon, W. K., and Chen, D.-R. (2005). Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. *Breast Cancer Research and Treatment*, 89(2):179–185.

Chatelain, P., Krupa, A., and Navab, N. (2015). Optimization of ultrasound image quality via visual servoing. In *Proc. of IEEE International Conference on Robotics and Automation*, pages 5997–6002.

Christensen, G. E. and Johnson, H. J. (2001). Consistent image registration. *Medical Imaging, IEEE Transactions on*, 20(7):568–582.

Cohen, B. and Dinstein, I. h. (2002). New maximum likelihood motion estimation schemes for noisy ultrasound images. *Pattern Recognition*, 35(2):455–463.

Comaniciu, D., Zhou, X., and Krishnan, S. (2004). Robust Real-Time Myocardial Border Tracking for Echocardiography: An Information Fusion Approach. *IEEE Transactions on Medical Imaging*, 23(7):849–860.

Coupé, P., Hellier, P., Morandi, X., and Barillot, C. (2012). 3d rigid registration of intraoperative ultrasound and preoperative MR brain images based on hyperechogenic structures. *Journal of Biomedical Imaging*, 2012:1.

De Luca, V., Székely, G., and Tanner, C. (2015). Estimation of Large-Scale Organ Motion in B-Mode Ultrasound Image Sequences: A Survey. *Ultrasound in Medicine & Biology*.

De Luca, V., Tschannen, M., Székely, G., and Tanner, C. (2013). A learning-based approach for fast and robust vessel tracking in long ultrasound sequences. In *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 518–525. Springer.

Delabarre, B. and Marchand, E. (2014). Dense non-rigid visual tracking with a robust similarity function. In *Proc. of IEEE International Conference on Image Processing*, pages 4942–4946.

Dietenbeck, T., Barbosa, D., Alessandrini, M., Jasaityte, R., Robesyn, V., D'hooge, J., Friboulet, D., and Bernard, O. (2014). Whole myocardium tracking in 2d-echocardiography in multiple orientations using a motion constrained level-set. *Medical Image Analysis*, 18(3):500–514.

Elen, A., Hon Fai Choi, Loeckx, D., Hang Gao, Claus, P., Suetens, P., Maes, F., and D'hooge, J. (2008). Three-Dimensional Cardiac Strain Estimation Using Spatio-Temporal Elastic Registration of Ultrasound Images: A Feasibility Study. *IEEE Transactions on Medical Imaging*, 27(11):1580–1591.

Fletcher, R. and Reeves, C. M. (1964). Function minimization by conjugate gradients. *The Computer Journal*, 7(2):149–154.

Gee, J. C. (1999). On matching brain volumes. *Pattern Recognition*, 32(1):99–111.

Gilles, M., Fayad, H., Miglierini, P., Clement, J. F., Scheib, S., Cozzi, L., Bert, J., Boussion, N., Schick, U., Pradier, O., and Visvikis, D. (2016). Patient positioning in radiotherapy based on surface imaging using time of flight cameras. *Medical Physics*, 43(8):4833–4841.

Gordon, N. J., Salmond, D. J., and Smith, A. F. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In *Proc. of International Conference on Radar and Signal Processing*, volume 140, pages 107–113.

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

Heimann, T., van Ginneken, B., Styner, M., Arzhaeva, Y., Aurich, V., Bauer, C., Beck, A., Becker, C., Beichel, R., Bekes, G., Bello, F., Binnig, G., Bischof, H., Bornik, A., Cashman, P., Chi, Y., Cordova, A., Dawant, B., Fidrich, M., Furst, J., Furukawa, D., Grenacher, L., Hornegger, J., Kainmuller, D., Kitney, R., Kobatake, H., Lamecker, H., Lange, T., Lee, J., Lennon, B., Li, R., Li, S., Meinzer, H.-P., Nemeth, G., Raicu, D., Rau, A.-M., van Rikxoort, E., Rousson, M., Rusko, L., Saddi, K., Schmidt, G., Seghers, D., Shimizu, A., Slagmolen, P., Sorantin, E., Soza, G., Susomboon, R., Waite, J., Wimmer, A., and Wolf, I. (2009). Comparison and Evaluation of Methods for Liver Segmentation From CT Datasets. *IEEE Transactions on Medical Imaging*, 28(8):1251–1265.

Hellier, P., Coupé, P., Morandi, X., and Collins, D. L. (2010). An automatic geometrical and statistical method to detect acoustic shadows in intraoperative ultrasound brain images. *Medical Image Analysis*, 14(2):195–204.

Heyde, B., Claus, P., Jasaityte, R., Barbosa, D., Bouchez, S., Vandenheuvel, M., Wouters, P., Maes, F., and Hooge, J. D. (2012). Motion and deformation estimation of cardiac ultrasound sequences using an anatomical B-spline transformation model. In *Proc. of IEEE International Symposium on Biomedical Imaging*, pages 266–269.

Higgins, H. and Berger, D. L. (2006). RFA for Liver Tumors: Does It Really Work? *The Oncologist*, 11(7):801–808.

Häme, Y., Gamarnik, V., Parker, K. M., Holmes, J. W., and Laine, A. F. (2012). Level set-based tracking of the endocardium without a shape prior from 3d ultrasound images. In *Proc. of IEEE International Symposium on Biomedical Imaging*, pages 466–469.

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

Hu, Y., Morgan, D., Ahmed, H. U., Pendsé, D., Sahu, M., Allen, C., Emberton, M., Hawkes, D., and Barratt, D. (2008). A statistical motion model based on biomechanical simulations for data fusion during image-guided prostate interventions. In *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 737–744. Springer.

Irani, M., Rousso, B., and Peleg, S. (1992). Detecting and tracking multiple moving objects using temporal integration. In Sandini, G., editor, *Proc. of European Conference on Computer Vision*, pages 282–287.

Jain, J. and Jain, A. (1981). Displacement measurement and its application in interframe image coding. *IEEE Transactions on communications*, 29(12):1799–1808.

Kadoury, S., Zagorchev, L., Wood, B. J., Venkatesan, A., Weese, J., Jago, J., and Kruecker, J. (2012). A model-based registration approach of preoperative MRI with 3d ultrasound of the liver for Interventional guidance procedures. In *Proc. of IEEE International Symposium on Biomedical Imaging*, pages 952–955.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45.

Karamalis, A., Wein, W., Klein, T., and Navab, N. (2012). Ultrasound confidence maps using random walks. *Medical Image Analysis*, 16(6):1101–1112.

Künzli, B. M., Abitabile, P., and Maurer, C. A. (2011). Radiofrequency ablation of liver tumors: Actual limitations and potential solutions in the future. *World Journal of Hepatology*, pages 8–14.

Kuklisova-Murgasova, M., Quaghebeur, G., Hajnal, J. V., Noble, J. A., and Schnabel, J. A. (2012). Towards 3d registration of fetal brain MRI and ultrasound. In *Proc. of IEEE International Symposium on Biomedical Imaging*, pages 346–349.

Lange, T., Papenberg, N., Heldmann, S., Modersitzki, J., Fischer, B., Lamecker, H., and Schlag, P. M. (2008). 3d ultrasound-CT registration of the liver using combined landmark-intensity information. *International Journal of Computer Assisted Radiology and Surgery*, 4(1):79–88.

Lasso, A., Heffter, T., Rankin, A., Pinter, C., Ungi, T., and Fichtinger, G. (2014). PLUS: Open-Source Toolkit for Ultrasound-Guided Intervention Systems. *IEEE Transactions on Biomedical Engineering*, 61(10):2527–2537.

Ledesma-Carbayo, M. J., Kybic, J., Desco, M., Santos, A., and Unser, M. (2001). Cardiac motion analysis from ultrasound sequences using non-rigid registration. In *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 889–896. Springer.

Lediju, M., Byram, B. C., Harris, E. J., Evans, P. M., Bamber, J. C., and others (2010). 3d Liver tracking using a matrix array: Implications for ultrasonic guidance of IMRT. In *Proc. of IEEE Ultrasonics Symposium*, pages 1628–1631.

Lee, D. and Krupa, A. (2011). Intensity-based visual servoing for non-rigid motion compensation of soft tissue structures due to physiological motion using 4d ultrasound. In *Proc. of IEEE International Conference on Intelligent Robots and Systems*, pages 2831–2836.

Lee, D., Nam, W. H., Lee, J. Y., and Ra, J. B. (2011). Non-rigid registration between 3d ultrasound and CT images of the liver based on intensity and gradient information. *Physics in Medicine and Biology*, 56(1):117–137.

Lubke, D. and Grozea, C. (2014). High Performance Online Motion Tracking in Abdominal Ultrasound Imaging. In *Proc. of MICCAI Workshop on Challenge on Liver Ultrasound Tracking*, page 29.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. In *Proc. of the National Institute of Sciences of India*, pages 49–56.

Marami, B., Sirouspour, S., Fenster, A., and W. Capson, D. (2014). Dynamic Tracking of a Deformable Tissue Based on 3d-2d MR-US Image Registration. In *Proc. of SPIE Medical Imaging*, pages 33–36.

Masum, M. A., Pickering, M., Lambert, A., Scarvell, J., and Smith, P. (2014). Accuracy assessment of Tri-plane B-mode ultrasound for non-invasive 3d kinematic analysis of knee joints. *Biomedical engineering online*, 13(1):122.

Metz, C., Klein, S., Schaap, M., van Walsum, T., and Niessen, W. (2011). Nonrigid registration of dynamic medical imaging data using nD+t B-splines and a groupwise optimization approach. *Medical Image Analysis*, 15(2):238–249.

Mikic, I., Krucinski, S., and Thomas, J. D. (1998). Segmentation and tracking in echocardiographic sequences: active contours guided by optical flow estimates. *IEEE Transaction on Medical Imaging*, 17(2):274–284.

Moré, J. J. and Thuente, D. J. (1994). Line Search Algorithms with Guaranteed Sufficient Decrease. *ACM Transactions on Mathematical Software*, 20(3):286–307.

Mukherjee, R., Sprouse, C., Abraham, T., Hoffmann, B., McVeigh, E., Yuh, D., and Burlina, P. (2011). Myocardial motion computation in 4d ultrasound. In *Proc. of IEEE International Symposium on Biomedical Imaging*, pages 1070–1073.

Nascimento, J. C. and Marques, J. S. (2008). Robust shape tracking with multiple models in ultrasound images. *IEEE Transactions on Image Processing*, 17(3):392–406.

Nazem, F., Ahmadian, A., Seraj, N. D., and Giti, M. (2013). Two-stage point-based registration method between ultrasound and CT imaging of the liver based on ICP and unscented Kalman filter: a phantom study. *International Journal of Computer Assisted Radiology and Surgery*.

Ni, D., Qu, Y., Yang, X., Chui, Y. P., Wong, T.-T., Ho, S. S., and Heng, P. A. (2008). Volumetric ultrasound panorama based on 3d SIFT. In *Proc. of Medical Image Computing and Computer-Assisted Intervention*, pages 52–60. Springer.

Nickolls, J., Buck, I., Garland, M., and Skadron, K. (2008). Scalable parallel programming with CUDA. *ACM Queue*, 6(2):40–53.

Orderud, F., Hansg\a ard, J., and Rabben, S. I. (2007). Real-time tracking of the left ventricle in 3d echocardiography using a state estimation approach. In *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 858–865. Springer.

Osher, S. and Fedkiw, R. P. (2003). *Level set methods and dynamic implicit surfaces*. Applied mathematical science. Springer, New York, N.Y.

Papademetris, X., Sinusas, A., Dione, D., Constable, R., and Duncan, J. (2002). Estimation of 3-D left ventricular deformation from medical images using biomechanical models. *IEEE Transactions on Medical Imaging*, 21(7):786–800.

Pennec, X., Cachier, P., and Ayache, N. (2001). Tracking brain deformations in time-sequences of 3d US images. In *Proc. of International Conference on Information Processing in Medical Imaging*, pages 169–175. Springer.

Penney, G., Blackall, J., Hamady, M., Sabharwal, T., Adam, A., and Hawkes, D. (2004). Registration of freehand 3d ultrasound and magnetic resonance liver images. *Medical Image Analysis*, 8(1):81–91.

Pickering, M. R., Muhit, A., Scarvell, J. M., Smith, P. N., and others (2009). A new multi-modal similarity measure for fast gradient-based 2d-3d image registration. In *Proc. of IEEE International Conference on Engineering in Medicine and Biology*, pages 5821–5824.

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

Richa, R., Souza, M., Scandaroli, G., Comunello, E., and von Wangenheim, A. (2014). Direct visual tracking under extreme illumination variations using the sum of conditional variance. In *Proc. of IEEE International Conference on Image Processing*, pages 373–377.

Richa, R., Sznitman, R., Taylor, R., and Hager, G. (2011). Visual tracking using the sum of conditional variance. In *Proc. of IEEE International Conference on Intelligent Robots and Systems*, pages 2953–2958.

Roche, A., Pennec, X., Malandain, G., and Ayache, N. (2001). Rigid registration of 3-D ultrasound with MR images: a new approach combining intensity and gradient information. *IEEE Transactions on Medical Imaging*, 20(10):1038–1049.

Rothlubbers, S., Schwaab, J., Jenne, J., and Gunther, M. (2014). MICCAI CLUST 2014-Bayesian Real-Time Liver Feature Ultrasound Tracking. In *Proc. of MICCAI Workshop on Challenge on Liver Ultrasound Tracking*, page 45.

Royer, L., Marchal, M., Le Bras, A., Dardenne, G., and Krupa, A. (2015). Real-time Tracking of Deformable Target in 3d Ultrasound Images. In *Proc. of IEEE International Conference on Robotics and Automation*.

Salles, S., Zahnd, G., Liebgott, H., Serusclat, A., and Vray, D. (2012). Real time US-tagging combined with phase-based optical flow applied to 2d motion estimation of the carotid artery wall. In *Proc. of IEEE Ultrasonics Symposium*, pages 1185–1188.

Schneider, R. J., Perrin, D. P., Vasilyev, N. V., Marx, G. R., del Nido, P. J., and Howe, R. D. (2012). Real-time image-based rigid registration of three-dimensional ultrasound. *Medical Image Analysis*, 16(2):402–414.

Schroeder, W., Martin, K. W., and Lorensen, B. (2002). *The visualization toolkit: an object-oriented approach to 3D graphics*. Kitware, Inc.

Sederberg, T. W. and Parry, S. R. (1986). Free-form deformation of solid geometric models. *ACM SIGGRAPH computer graphics*, 20(4):151–160.

Seo, J., Koizumi, N., Funamoto, T., Sugita, N., Yoshinaka, K., Nomiya, A., Homma, Y., Matsumoto, Y., and Mitsuishi, M. (2011). Visual servoing for a US-guided therapeutic HIFU system by coagulated lesion tracking: a phantom study. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 7(2):237–247.

Shekhar, R. and Zagrodsky, V. (2002). Mutual information-based rigid and nonrigid registration of ultrasound volumes. *IEEE Transactions on Medical Imaging*, 21(1):9–22.

Si, H. (2015). TetGen, a Delaunay-Based Quality Tetrahedral Mesh Generator. *ACM Transaction on Mathematical Software*, 41(11).

Somphone, O., Allaire, S., Mory, B., and Dufour, C. (2014). Live Feature Tracking in Ultrasound Liver Sequences with Sparse Demons. In *Proc. of MICCAI Workshop on Challenge on Liver Ultrasound Tracking*, page 53.

Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190.

Tanner, C., Schnabel, J. A., Chung, D., Clarkson, M. J., Rueckert, D., Hill, D. L., and Hawkes, D. J. (2000). Volume and shape preservation of enhancing lesions when applying non-rigid registration to a time series of contrast enhancing MR breast images. In *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 327–337. Springer.

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012: Global Cancer Statistics, 2012. *CA: A Cancer Journal for Clinicians*, 65(2):87–108.

Touil, B., Basarab, A., Delachartre, P., Bernard, O., and Friboulet, D. (2010). Analysis of motion tracking in echocardiographic image sequences: Influence of system geometry and point-spread function. *Ultrasonics*, 50(3):373–386.

Uhercik, M., Liebgott, H., Kybic, J., and Cachard, C. (2009). Needle localization methods in 3d ultrasound data. In *Proc. of International Congress on Ultrasonics*, page 60.

Veronesi, F., Corsi, C., Caiani, E. G., and Lamberti, C. (2005). Nearly automated left ventricular long axis tracking on real time three-dimensional echocardiographic data. In *Proc. of IEEE International Conference on Computers in Cardiology*, pages 5–8.

Vijayan, S., Klein, S., Hofstad, E. F., Lindseth, F., Ystgaard, B., and Lango, T. (2013). Validation of a non-rigid registration method for motion compensation in 4d ultrasound of the liver. In *Proc. of IEEE International Symposium on Biomedical Imaging*, pages 792–795.

Wei, Z., Ding, M., Downey, D., and Fenster, A. (2005). 3d TRUS guided robot assisted prostate brachytherapy. In *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 17–24. Springer.

Wein, W., Brunke, S., Khamene, A., Callstrom, M. R., and Navab, N. (2008a). Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. *Medical Image Analysis*, 12(5):577–585.

Wein, W., Cheng, J.-Z., and Khamene, A. (2008b). Ultrasound based respiratory motion compensation in the abdomen. In *Proc. of MICCAI Worshop on Image Guidance and Computer Assistance for Soft tissue Interventions*, volume 32, page 294.

Wein, W., Kutter, O., Aichert, A., Zikic, D., Kamen, A., and Navab, N. (2010). Automatic non-linear mapping of pre-procedure CT volumes to 3d ultrasound. In *Proc. of IEEE International Symposium on Biomedical Imaging*, pages 1225–1228.

Xiao, D., Ng, W. S., Tsang, C. B., and Abeyratne, U. R. (2007). A region and gradient based active contour model and its application in boundary tracking on anal canal ultrasound images. *Pattern Recognition*, 40(12):3522–3539.

Yeung, F., Levinson, S. F., Fu, D., and Parker, K. J. (1998). Feature-adaptive motion tracking of ultrasound image sequences using a deformable mesh. *IEEE Transaction on Medical Imaging*, 17(6):945–956.

Yipeng Hu, Carter, T. J., Ahmed, H. U., Emberton, M., Allen, C., Hawkes, D. J., and Barratt, D. C. (2011). Modelling Prostate Motion for Data Fusion During Image-Guided Interventions. *IEEE Transactions on Medical Imaging*, 30(11):1887–1900.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., and Gerig, G. (2006). User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, 31(3):1116–1128.

Zhang, X., Günther, M., and Bongers, A. (2010). Real-time organ tracking in ultrasound imaging using active contours and conditional density propagation. In *Proc. of International Conference on Medical Imaging and Augmented Reality*, pages 286–294. Springer.

Appendix A

Publications and patents

A.1 Scientific journal

L. Royer, A. Krupa, G. Dardenne, A. Le Bras, E. Marchand, M. Marchal. Real-time Target Tracking of Soft Tissues in 3D Ultrasound Images Based on Robust Visual Information and Mechanical Simulation. *Medical Image Analysis*, accepted, available online

A.2 International Conferences

L. Royer, M. Marchal, A. Le Bras, G. Dardenne, A. Krupa. Real-time tracking of deformable target in 3D ultrasound images. In *Proceedings of IEEE International Conference on Robotics and Automation*, ICRA'15, Pages 2430-2435, Seattle, WA, Mai 2015

L. Royer, G. Dardenne, A. Le Bras, M. Marchal, A. Krupa. Tracking of Non-rigid Targets in 3D US Images: Results on CLUST 2015. In *Proceedings of MICCAI 2015 Challenge on Liver Ultrasound Tracking*, Munich, Germany, Octobre 2015

L. Royer, M. Marchal, A. Le Bras, G. Dardenne, A. Krupa. Tracking of Deformable Target in 2D Ultrasound Images. In *Proceedings of SPIE Medical Imaging Conference*, Orlando, USA, Février 2015

L. Royer, M. Marchal, A. Le Bras, G. Dardenne, A. Krupa. Interactive Tracking of Soft Tissues in 2D Ultrasound Images. In *Proceedings of SURGETICA*, 2014, Chambery, France, Décembre 2014

L. Royer, M. Babel, A. Krupa. *Non-rigid target tracking in 2D ultrasound images using hierarchical grid interpolation.* In *Proceedings of SPIE Medical Imaging Conference*, San Diego, CA, Février 2014

A.3 Patent

L. Royer, M. Marchal, A. Krupa. *Process for tracking a clinical target in medical images*, France, FR15/60541, 2015

Appendix B

External displacement computation

In this section, details are given about the computation of the optimal displacement parameters. Concerning the cost function expressed in Eq. 4.8, we have:

$$C(\Delta\mathbf{q}) \approx (\mathbf{DJ}\Delta\mathbf{q} + \mathbf{D}(I_t(\mathbf{Mq}_t^{k-1}) - I_{t_0}^*(\mathbf{Mq}_{t_0})))^2 \tag{B.1}$$

The optimal parameters are obtained by minimizing the cost function C such that we obtain the derivative of C with respect to $\Delta\mathbf{q}$ equal to zero:

$$\frac{\delta C(\Delta\mathbf{q})}{\delta\Delta\mathbf{q}} = 0 \tag{B.2}$$

It follows that:

$$\frac{\delta C(\Delta\mathbf{q})}{\delta\Delta\mathbf{q}} = 2(\mathbf{DJ})^T(\mathbf{DJ}\Delta\mathbf{q} + \mathbf{D}(I_t(\mathbf{Mq}_t^{k-1}) - I_{t_0}^*(\mathbf{Mq}_{t_0}))) = 0 \tag{B.3}$$

In order to compute the optimal parameters, we can rewrite the previous expression such that:

$$\Delta\mathbf{q} = (\mathbf{J}^T\mathbf{D}^T\mathbf{DJ})^{-1}\mathbf{J}^T\mathbf{D}^T\mathbf{D}(I_t(\mathbf{Mq}_t^{k-1}) - I_{t_0}^*(\mathbf{Mq}_{t_0})) \tag{B.4}$$

$$\Delta\mathbf{q} = -(\mathbf{DJ})^+\mathbf{D}(I_t(\mathbf{Mq}_t^{k-1}) - I_{t_0}^*(\mathbf{Mq}_{t_0})) \tag{B.5}$$

where $(\mathbf{DJ})^+$ represents the left pseudo-inverse of the matrice $\mathbf{DJ}$. However, the computation of the pseudo-inverse is highly computational demanding due to its size $(N_v) \times (3 \cdot N_c)$. In our practical experiments, the size of the Jacobian matrix may be up to 20000×600 and its pseudo-inverse computation may require up to 2.5 seconds per iteration. To cope with that issue, we propose to use the steepest gradient strategy that consists in replacing the inverse

term such that:

$$\mathbf{J}^T\mathbf{D}^T\mathbf{D}\mathbf{J} \approx \mathbf{J}^T\mathbf{D}^T\mathbf{D}\mathbf{J} + \lambda\mathbf{I} \tag{B.6}$$

where $\mathbf{I}$ denotes the identify matrix. λ represent a damping coefficient. By assuming that the coefficient λ is high enough, we can ignore the left term of the previous equation. We therefore obtain:

$$\Delta\mathbf{q} = -\alpha\mathbf{J}^T\mathbf{D}^T\mathbf{D}[I_t(\mathbf{M}(\mathbf{q}^{k-1}(t))) - I^*_{t_0}(\mathbf{M}(\mathbf{q}(t_0)))] \tag{B.7}$$

INSA de RENNES
Service des Formations

AVIS DU JURY SUR LA REPRODUCTION DE LA THESE SOUTENUE

Titre de la thèse:
Real-time tracking of deformable targets in 3D ultrasound sequences

Nom Prénom de l'auteur : ROYER LUCAS

Membres du jury :
- Monsieur KRUPA Alexandre
- Madame MARCHAL Maud
- Madame TROCCAZ Jocelyne
- Monsieur LIEBGOTT Hervé
- Monsieur COTIN Stéphane
- Monsieur DARDENNE Guillaume

Président du jury : Stéphane COTIN

Date de la soutenance : 06 Décembre 2016

Reproduction de la these soutenue

☒ Thèse pouvant être reproduite en l'état
~~Thèse pouvant être reproduite après corrections suggérées~~

Fait à Rennes, le 06 Décembre 2016

Signature du président de jury

Le Directeur,

M'hamed DRISSI

Résumé

De nos jours, les traitements mini-invasifs, tels que l'ablation par radiofréquence, sont de plus en plus utilisés car ils permettent d'éliminer localement les tumeurs à partir de l'insertion d'une aiguille. Cependant, le succès de ces procédures dépend de la précision du positionnement de l'aiguille par rapport aux structures anatomiques. Afin de garantir un placement correct, l'imagerie échographique est souvent utilisée car elle a l'avantage d'être temps-réel, bas coût, et non-invasive. En revanche, cette modalité peut compliquer la visualisation de certaines structures en raison de sa qualité et de son champ de vue limité. En outre, la précision des interventions peut aussi être perturbée par les déplacements de tissus liés aux mouvements physiologiques du patient et à la manipulation d'instruments médicaux. Afin d'aider le chirurgien à mieux cibler certaines structures anatomiques, de nombreuses équipes de recherche ont proposé des travaux permettant d'estimer la position de régions d'intérêts dans l'imagerie échographique.

Cette thèse propose plusieurs contributions permettant de suivre en temps-réel des structures déformables dans des séquences d'échographie 3D. Une première contribution repose sur l'utilisation conjointe de l'information visuelle dense et d'une méthode de simulation physique. Dans cette thèse, nous avons aussi proposé un nouveau critère de similarité spécifique à l'imagerie échographique basé sur une étape de détection d'ombres. Enfin, la dernière contribution est liée à une stratégie de suivi hybride permettant d'améliorer la qualité des images. A partir de ces contributions, nous proposons une méthode de suivi robuste au bruit de type « speckle », aux ombres et aux changements d'intensité perturbant l'imagerie échographique.

Les performances des différentes contributions sont évaluées à partir de données simulées et de données acquises sur maquettes et sur volontaires humains. Ces résultats montrent que notre méthode est robuste à différents artefacts de l'imagerie échographique. En outre, nous démontrons la performance de notre approche par rapport à différentes méthodes de l'état de l'art sur des bases de données publiques fournies par les challenges MICCAI CLUST'14 et CLUST'15.

Dans cette thèse, nous proposons également une application permettant de combiner l'imagerie échographique à l'imagerie par résonance magnétique (IRM). Cette méthode permet d'observer des structures anatomiques non-visibles dans l'imagerie échographique durant l'intervention. Elle est basée sur la combinaison d'une méthode de suivi et d'un recalage multi-modal obtenu à partir d'un système de localisation externe. Cette application a été évaluée sur un volontaire sain à partir d'une plateforme liée au centre Hospitalier Universitaire de Rennes.

Abstract

Nowadays, mini-invasive treatments, such as radio-frequency ablation, are increasingly being used because they allow eliminating tumors locally from needle insertion. However, the success of these therapies depends on the accurate positioning of the needle with respect to anatomical structures. To ensure correct placement, ultrasound (US) imaging is often used since this system has the advantage to be real-time, low-cost, and non-invasive. However, during the intervention, US imaging can complicate the visualization of targeted structures due to its poor quality and its limited field of view. Furthermore, the accuracy of these interventions may also be perturbed by both physiological movements and medical tools displacements that introduce motions of anatomical structures. To help the surgeon to better target malignant tissues, many research teams have proposed different methods in order to estimate the position of regions of interest in ultrasound imaging.

This thesis provides several contributions that allow tracking deformable structures in 3D ultrasound sequences. We first present a method that allows providing robust estimation of target positions by combining an intensity-based approach and mechanical model simulation. In this thesis, we also propose novel ultrasound-specific similarity criterion based on prior step that aims at detecting shadows. The last contribution is related to a hybrid tracking strategy that allows improving quality of ultrasound images. From these contributions, we propose a tracking method that has the advantage to be invariant to speckle noise, shadowing and intensity changes that can occur in US imaging.

The performance and limitations of the proposed contributions are evaluated through simulated data, phantom data, and real-data obtained from different volunteers. Simulation and phantom results show that our method is robust to several artefacts of US imaging such as shadows and speckle decorrelation. Furthermore, we demonstrate that our approach outperforms state-of-the-art methods on the 3D public databases provided by MICCAI CLUST'14 and CLUST'15 challenges.

In this thesis, we also propose an application that combines ultrasound imaging to Magnetic Resonance Imaging (MRI). This method allows observing anatomical structures that are not visible in US imaging during the intervention. It is based on the combination between US tracking method and multi-modal registration obtained from external localization system. This application was evaluated on a volunteer thanks to an MRI imaging platform located at the University Hospital of Rennes.

INSA RENNES

N° d'ordre : **16ISAR 23 / D16 - 23**

Institut National des Sciences Appliquées de Rennes
20, Avenue des Buttes de Coësmes - CS 14315 - F-35043 Rennes Cedex
Tél : 02 23 23 82 00 – Fax : 02 23 23 83 96