



**HAL**  
open science

# Real-time Target Tracking of Soft Tissues in 3D Ultrasound Images Based on Robust Visual Information and Mechanical Simulation

Lucas Royer, Alexandre Krupa, Guillaume Dardenne, Anthony Le Bras, Eric  
Marchand, Maud Marchal

► **To cite this version:**

Lucas Royer, Alexandre Krupa, Guillaume Dardenne, Anthony Le Bras, Eric Marchand, et al.. Real-time Target Tracking of Soft Tissues in 3D Ultrasound Images Based on Robust Visual Information and Mechanical Simulation. *Medical Image Analysis*, 2017, 35, pp.582 - 598. 10.1016/j.media.2016.09.004 . hal-01374589

**HAL Id: hal-01374589**

**<https://inria.hal.science/hal-01374589v1>**

Submitted on 30 Sep 2016

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

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

# Real-time Target Tracking of Soft Tissues in 3D Ultrasound Images Based on Robust Visual Information and Mechanical Simulation

Lucas Royer<sup>a,b,c</sup>, Alexandre Krupa<sup>a,b</sup>, Guillaume Dardenne<sup>a</sup>, Anthony Le  
Bras<sup>a,d</sup>, Eric Marchand<sup>e,f</sup>, Maud Marchal<sup>a,c,f</sup>

<sup>a</sup>*Institut de Recherche Technologique b-com, Rennes, France*

<sup>b</sup>*Inria Rennes - Bretagne Atlantique, France*

<sup>c</sup>*INSA de Rennes, France*

<sup>d</sup>*CHU de Rennes, France*

<sup>e</sup>*Université de Rennes 1, France*

<sup>f</sup>*IRISA, Rennes, France*

---

## Abstract

In this paper, 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 robust dense motion estimation and mechanical model simulation. We perform evaluation of our method through simulated data, phantom data, and real-data. Results demonstrate that this novel approach has the advantage of providing correct motion estimation regarding different ultrasound shortcomings including speckle noise, large shadows and ultrasound gain variation. 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.

*Keywords:* 3D ultrasound images, Tracking, Dense information, Deformable model

---

## 1. Introduction

Soft-tissue motion tracking is an active research area that consists in providing accurate evaluation about the location of anatomical structures. To do so, ultrasound imaging is often used since it is non-invasive, real-time and portable. Thus, several ultrasound tracking approaches have been developed in order to

estimate soft tissue displacements that are caused by physiological motions and manipulations by medical tools. These methods have gained significant interest for image-guided therapies such as radio-frequency ablation (RFA) (Higgins and Berger, 2006) or high-intensity focused ultrasound (HIFU) (Pernot et al., 2004) that consist in eliminating tumors by delivering a local treatment on a targeted anatomical region. However, these tracking techniques remain sensitive to different ultrasound imaging shortcomings such as large ultrasound shadows, gain change and speckle noise. In this paper, we propose a novel tracking approach to tackle these limitations. Our method combines an intensity-based approach with a mechanical regularization. We also propose an ultrasound-specific similarity criterion that has the advantage to be computationally efficient and robust to gain changes introduced by ultrasound imaging.

## 2. Related work

Different approaches have been proposed in the literature to track in real-time a clinical target. A first category of methods consists in aligning key features over a sequence of US images sequence such as surfaces (Papademetris et al., 2002), or keypoints including scale-invariant feature transform (SIFT) descriptors (Schneider et al., 2012). To improve robustness against noise, these methods can be based on a Bayesian framework in order to include prior knowledge about the target shape over the time (Angelova and Mihaylova, 2010; Rothlubbers et al., 2014; Zhang et al., 2010). However, such approaches might fail when key features are not visible due to the presence of large shadows and other ultrasound artifacts. Another category of methods is based on the minimization of an intensity cost function by using mono-modal similarity metric such as Sum of Squared Difference (SSD) (Lubke and Grozea, 2014; Royer et al., 2015; Yeung et al., 1998), Sum of Absolute Difference (SAD) (Touil et al., 2010), or Cross-Correlation (CC) (Basarab et al., 2008; De Luca et al., 2013). These criteria are generally well-suited if the target intensity remains constant over the time. However, in ultrasound-guided procedures, this assumption may be

easily violated and may produce inaccurate motion estimation. To cope with the previous issue, ultrasound-specific similarity measures have been proposed (Cohen and Dinstein, 2002; Baumann et al., 2012). Baumann et al. (2012) proposed a correlation-based distance measure that is able to deal 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. Cohen and Dinstein (2002) proposed a similarity measure that assumes that the US images are degraded by Raleigh distribution multiplicative noise. This is therefore well suited to high speckle noise but is inefficient when large shadows occur. Multi-modal criteria have also been studied such as Mutual Information (MI) (Elen et al., 2008; Shekhar and Zagrodsky, 2002), or Sum of Conditional Variance (SCV) (Masum et al., 2014) since they are more robust to ultrasound gain variations. However, MI calculation is computationally expensive and is therefore not well suited for applications of real-time tracking in a sequence of 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.

Ultrasound tracking methods can also be classified regarding their warping models that allows defining the possible spatial transformation. Several warping functions have been studied such as translational model (Veronesi et al., 2005), or affine model (Wein et al., 2008). However, these warping models are not well-suited for deformable structures. To cope with that issue, deformable transformation models can also be used. Some of them rely on the classical block matching algorithm that estimates the target motions by computing the displacement of small image blocks between consecutive frames (Basarab et al., 2008; De Luca et al., 2013; Touil et al., 2010; Yeung et al., 1998). However, the block-matching methods can not represent highly localized deformable motions since they assume that the displacement is rigid within a local region block. To solve this issue, approaches have been proposed to estimate the dense motion field from deformable models such as piece-wise affine model (Royer et al., 2015), thin-plate spline model (Lee and Krupa, 2011), or free-form deformation

(Heyde et al., 2012; Ledesma-Carbayo et al., 2001; Pennec et al., 2001). To ensure robustness, these methods can be combined to spatio-temporal smoothness constraint (Pennec et al., 2001; Somphone et al., 2014) or coarse-to-fine optimization (Mukherjee et al., 2011). Furthermore, other extensions improving the accuracy can consist in registering the ultrasound images in both forward and backward direction (Ledesma-Carbayo et al., 2001), or in a group-wise optimization fashion (Metz et al., 2011; Vijayan et al., 2013), or by using specific grid adapted to the target instead of regular rectangular grid (Heyde et al., 2012). An outlier rejection can also be coupled to these techniques in order to achieve more accurate results regarding speckle noise (Banerjee et al., 2015). In addition to these techniques that are only based on visual criterion optimization, mechanical-based tracking approaches have also been proposed for 2D ultrasound images (Loosvelt et al., 2014; Marami et al., 2014; Yeung et al., 1998). Nevertheless, these methods are not well-suited during out-of-plane motions since they can cause non-physically plausible displacements. To cope with that issue, Yipeng Hu et al. (2011) proposed a 3D tracking method that is regularized by statistical motion model obtained from biomechanical modelling. However, their method requires identifying manually some surfaces points of the prostate over each ultrasound frame in order to drive the model. To summarize, we present a classification of tracking methods according to their main features in table 1. To the best of our knowledge, no real-time tracking method combining robust dense method and mechanical model has been designed for 3D ultrasound images.

Method	2D/3D	Registration type			Manual Inter.	Pretreatment	Comp. Time
		Match. Crit.	Transformation	Regularization			
Papadimitris et al. (2002)	3D	Shape-Tracking	Gaussian	Mechanical	Surface Segm.	x	3 hours
Schneider et al. (2012)	3D	LoG	Rigid	Rausse Filtering	x	x	100 ms
Lubke and Groza (2014)	3D	SSD	Quadratic Poly.	Outlier rejection/Spatial	Histogram Equalization/Bilateral Filtering	x	61 ms
Toail et al. (2010)	2D	SAD	BM(Po)	Multiscale	x	x	x
Beserb 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
Shakhar and Zagrodsky (2002)	3D	MI	Affine	x	x	x	1 min
Masun et al. (2014)	2D	SCV	Rigid	x	x	x	x
Veronesi et al. (2005)	3D	SSD	Translational	Temporal	x	x	1 min
Wein et al. (2008)	2D	LNCC	Affine	Mechanical	x	x	2/10 ms
Young et al. (1998)	2D	SSD	BM	Temporal	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
Pennee et al. (2001)	3D	SSD/LCC	Gaussian	Spatial/Temporal	x	x	10 min
Somphone et al. (2014)	3D	SSD	Gaussian	Fluid operator	Dark Filtering	x	100 ms
Mukherjee et al. (2011)	3D	SSD	FFD	Multiscale/Spatial/Temporal	x	x	x
Baurjee et al. (2015)	3D	NCC	BM	Outlier rejection (GeoAppCons)	x	x	125 ms
Meitz 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
Marani et al. (2014)	2D	MIND	BM	FEM simulation	x	x	380 ms
Bell et al. (2012)	3D	NCC	Translational	x	Median Filtering	x	x
Ledja et al. (2010)	3D	NCC	Translational	x	Median Filtering	x	x
Baumann et al. (2012)	3D	CC / SC	VF	Inverse Consistency/Mechanical	Surface Pts Ident.	x	7 s
Yipeng Hu et al. (2011)	3D	SiP	FEM	Statistical Motion Model (FEA)	x	x	No
Angolova and Mihaylova (2010)	2D	MMPF	Deformable Contour	Spatial	Non-linear Gaussian filtering	x	4 min
Zhang et al. (2010)	2D	PDF	Deformable Contour	x	Surface Pts Ident.	x	3.4 ms
Rothhitters et al. (2014)	3D	PDF	Rigid	x	x	x	372 ms
Milic et al. (1998)	2D	SSD	Deformable Contour	Mechanical/Outlier rejection(VoFill)	x	x	6 s
Proposed method	3D	SCCV	Piecewise affine	Mechanical/Confidence maps	x	x	350 ms

Table 1: Match. Crit.: Matching criterion, (SSD) Sum of Squared Differences, (SAD) Sum of Absolute Differences, (CC) Cross-Correlation, (SC) Shift-Correlation (US-ML) Ultrasound Specific Maximum Likelihood estimator, (MIND) Modality Independent Neighborhood Descriptor, (PDF) Probability density function, (SFP) Surface to Points distance, (Learning) Learning-based method. (Av) Reference image intensity value 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 adaptive, (VF) Variation framework, (Quad. Poly.) Quadratic polynomial. Registration (Spatial/Temporal) Spatial/temporal smoothness, (Grpwise, VoFill) Velocity filtering, (GeoAppCons) Groupwise 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. Pre-treatment (TPV/DI) Trilinear Partial Volume Distribution Interpolation, (Dark filtering) Points in the darkest regions are discarded. (Quadratic Poly.) Quadratic polynomial, (x) Not specified. Comp. time Computation Time, (x) Not specified.

### 3. Method and Materials

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 3.1). 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 combined to a shadow detection process. The computational flow of the method is summarized in Fig. 1. As it can be seen from Fig. 1, the vertex

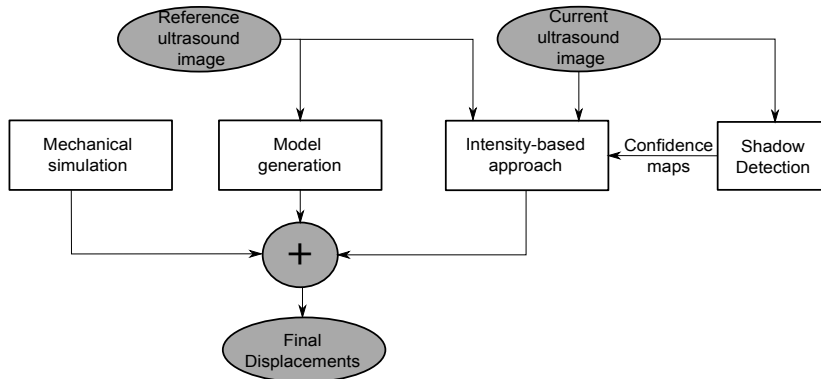


Figure 1: Computational flow of the method. The different steps of the method are represented by white squares. The input and output data are characterized by grey ellipses.

positions  $\mathbf{q}_k(t)$  of the model are computed by using the following equation:

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

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 3.2 and 3.3.  $\mathbf{q}_{k-1}(t)$  denotes the estimation of the previous vertex

positions at time index  $t$  at iteration  $k-1$  of the optimization algorithm.  $h_i$  represents a gain that amplifies the effect of external displacements estimation. It is similar to a weight that balances the contribution between external forces and internal forces obtained from mechanical model.

### 3.1. 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 a 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 as described by 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 piecewise affine warp function that is parameterized from both the vertex 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:

$$\mathbf{p} = \mathbf{M} \cdot \mathbf{q} \tag{2}$$

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 vertex positions. It is worth mentioning that  $\mathbf{p}$  represents only the voxels in the mesh model. Thanks to Eq. (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 vertex positions in US images, we combine a mechanical model to the estimation of displacement. The model generation step is illustrated in Fig. 2.



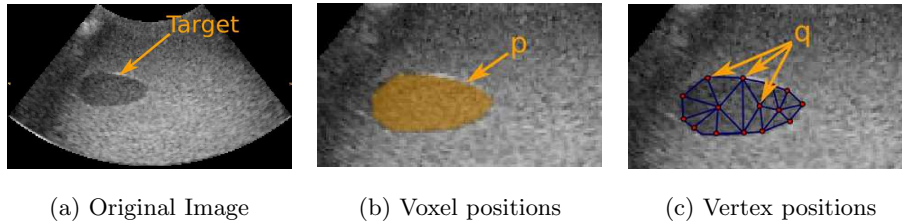


Figure 2: Illustration of the model generation step. (a) A anatomical region is identified in the first frame. (b-c) The associated voxel positions  $\mathbf{p}$  and the vertex positions  $\mathbf{q}$  are shown. For sake of clarity, tetrahedral cells are represented by triangular cells in this 2D illustration.

### 3.2. Estimation of Internal Displacements

Our approach combines a mechanical model based on mass-spring-damper system to the mesh previously described. Thus, the vertex 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 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:

$$\mathbf{f}_{i,j} = K_{ij}(d_{ij} - d_{ij}^{init})(\mathbf{q}_i - \mathbf{q}_j) + D_{ij}(\dot{\mathbf{q}}_i - \dot{\mathbf{q}}_j) \circ (\mathbf{q}_i - \mathbf{q}_j) \quad (3)$$

where  $d_{ij}$  and  $d_{ij}^{init}$  respectively represent the distance 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:

$$\mathbf{f}_i = \sum_{r=0}^{N_i} \mathbf{f}_{i,r} + G_i \dot{\mathbf{q}}_i \quad (4)$$

$N_i$  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) with a semi-implicit Euler integration scheme such that:

$$\Delta \mathbf{d} = \Delta_t (\dot{\mathbf{q}} + \Delta_t \mathbf{f}^{int}) \quad (5)$$

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

### 3.3. 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:

$$\hat{\mathbf{q}} = \arg \min_{\mathbf{q}} C(\mathbf{q}) = \arg \min_{\mathbf{q}} E(I_t(\mathbf{p}(t)), I_{t_0}(\mathbf{p}(t_0))) \quad (6)$$

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  $\mathbf{q}$ .  $\mathbf{p}(t_i)$  represents the voxel positions at time index  $t_i$ .  $E$  expresses the dissimilarity measure. In this work, we evaluated several dissimilarity functions including Sum of Squared Differences (SSD), Weighted Sum of Squared Differences (WSSD), Sum of Conditional Variance (SCV) and the proposed Sum of Confident Conditional Variance (SCCV).

#### 3.3.1. Sum of Squared Differences (SSD)

In US tracking applications based on dense methods, the Sum of Squared Differences (SSD) criterion is the most used dissimilarity measure since it has

the advantage of being less expensive than other criteria. The SSD cost function can be expressed as follows:

$$C(\mathbf{q}) = (I_t(\mathbf{p}(t)) - I_{t_0}(\mathbf{p}(t_0)))^2 \quad (7)$$

The methods based on SSD generally assume that the ultrasound wave reflected by a physical point is constant and time independent. However, in ultrasound-guided procedures, this assumption is not always valid due to the presence of shadows. To cope with this issue, we investigated new criteria able to integrate robust shadow detection process. It is also worth mentioning that speckle decorrelation may also perturb the SSD cost function since it introduces intensity variation. However, such issue may be solved by using adapted regularization term as shown in following sections and experiments proposed by Royer et al. (2015).

### 3.3.2. *Weighted Sum of Squared Differences (WSSD)*

Several automatic methods have recently emerged in order to detect shadows (Penney et al., 2004; Hellier et al., 2010; Karamalis et al., 2012). Penney et al. (2004) presented a shadow removal method that detects shadow along ultrasound scanline by removing low intensity region. Hellier et al. (2010) proposed an approach where shadowed regions are identified 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. 3. For each voxel, the range of confidence values can vary between the lowest to the highest confidence level. Once the confidence image  $U_t$  is defined, we propose to design

a robust weighted SSD cost function that we express as follows:

$$C(\mathbf{q}) = (\mathbf{H}_t(I_t(\mathbf{p}(t)) - I_{t_0}(\mathbf{p}(t_0))))^2 \quad (8)$$

where  $\mathbf{H}_t$  is a  $(N_v \times N_v)$  diagonal matrix computed from confidence image (map)  $U_t$ . This matrix allows respectively to discriminate unconfident voxels, that have confident value lower than a certain threshold  $\tau$ , and emphasize the others. From Eq. 8, we can notice that the weighting function is not applied to the reference image since we assume that its quality should be high enough. This ensures that the number of voxels is sufficient in order to obtain good tracking performance. The weighting coefficient  $\mathbf{H}_t(\mathbf{p}_k(t))$  of each voxel  $\mathbf{p}_k(t)$  is computed as follows:

$$\mathbf{H}_t(\mathbf{p}_k(t)) = \begin{cases} \frac{U_t(\mathbf{p}_k(t))^\beta}{\tau^\beta}, & \text{if } 0 \leq U_t(\mathbf{p}_k(t)) < \tau \\ 1, & \text{otherwise} \end{cases} \quad (9)$$

where  $\tau$  denotes the minimum confidence threshold that ensures the maximum weight ( $\mathbf{H}_t(\mathbf{p}_k(t)) = 1$ ).  $\beta$  is a scalar parameter that determine the smoothness of the weight matrix function. The evolution of this weighting function shape is illustrated in Fig. 4.

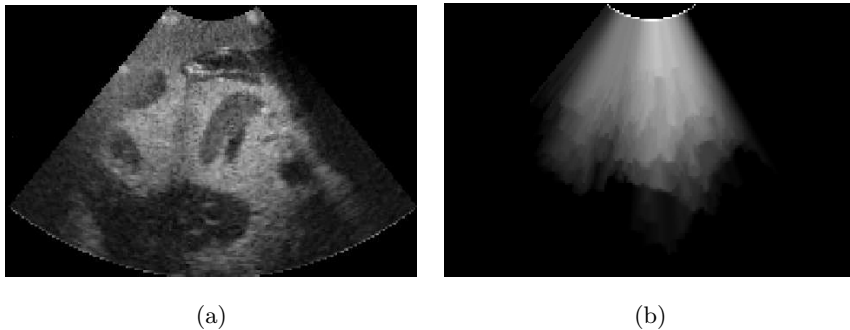


Figure 3: Illustration of the ultrasound image  $I_t$  (left) and its confidence image  $U_t$  (right).

### 3.3.3. Sum of Conditional Variance (SCV)

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

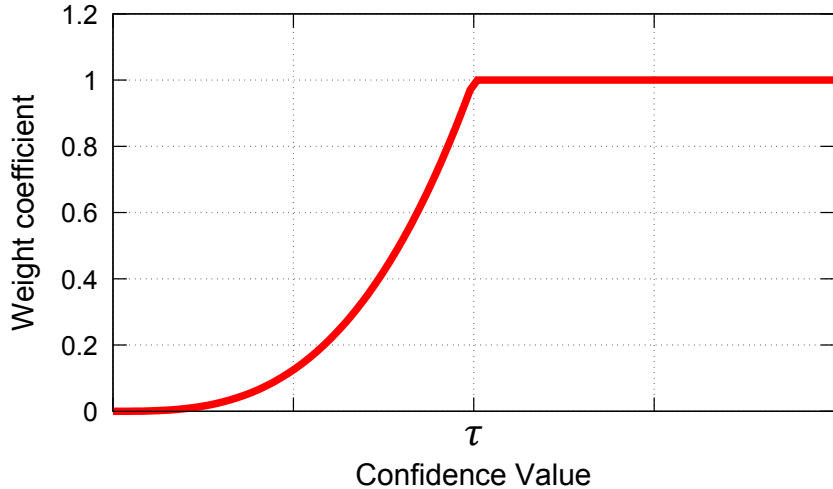


Figure 4: Weighting function that provides the weighting coefficient  $\mathbf{H}_t(\mathbf{p}_k(t))$  of a voxel  $\mathbf{p}_k(t)$  from its confidence value  $U_t(\mathbf{p}_k(t))$ .

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:

$$C(\mathbf{q}) = (I_t(\mathbf{p}(t)) - \widehat{I}_{t_0}^t(\mathbf{p}(t_0)))^2 \quad (10)$$

where  $\widehat{I}_{t_0}^t$  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:

$$\widehat{I}_{t_0}^t = \mathcal{E}(I_t|I_{t_0}) \quad (11)$$

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

$$\widehat{I}_{t_0}^t(\mathbf{p}_x(t_0)) = \sum_{i=0}^{L-1} i \frac{p_{I_t I_{t_0}}(i, j)}{p_{I_t}(j)} \quad \text{if } I_{t_0}(\mathbf{p}_x) = j \quad (12)$$

where  $L$  represents the number of grey level 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:

$$p_{I_t}(j) = \sum_{i=0}^{L-1} p_{I_t I_{t_0}}(i, j) \quad (13)$$

$$p_{I_t I_{t_0}}(i, j) = \sum_{k=1}^{N_v} \delta_i(I_t(\mathbf{p}_k(t))) \delta_j(I_{t_0}(\mathbf{p}_k(t_0))) \quad (14)$$

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

$$\delta_u(x) = \begin{cases} 1, & \text{if } x=u \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

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 paper a novel ultrasound-specific version of this criterion by using shadow detection.

#### 3.3.4. Sum of Confident Conditional Variance (SCCV)

To limit the effect of shadowed voxels, we propose to improve the previous criterion by modifying the cost function expressed in Eq. (10). To do so, we can refer to weighted cost function of SSD (Eq. (8)). The cost function can now be expressed as:

$$C(\mathbf{q}) = (\mathbf{H}_t(I_t(\mathbf{p}(t)) - \widehat{I}_{t_0}^t(\mathbf{p}(t_0))))^2 \quad (16)$$

However, in the original version of the SCV, the joint probability density function is computed from the intensities of all voxels positions of the model. However, due to the presence of shadows, these intensities may partially change on the location of shadowed voxels. This produces an inaccurate estimation of the probability density function  $p_{I_t I_{t_0}}(i, j)$ . Consequently, the new intensities of the adapted reference frame can be mis-estimated. 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:

$$p_{I_t I_{t_0}}(i, j) = \sum_{k=1}^{N_v} \mathbf{H}_t(\mathbf{p}_k(t)) \delta_i(I_t(\mathbf{p}_k(t))) \delta_j(I_{t_0}(\mathbf{p}_k(t_0))) \quad (17)$$

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. (9) from the confidence map. Thus, the computation of the adapted reference frame  $\widehat{I}_{t_0}^t$  is not disrupted by voxels that are located in shadowed regions. By doing these modifications, we propose an ultrasound-specific dissimilarity measure that has the advantage to be robust to global and local intensity changes caused by shadows presence or gain variation of the ultrasound imaging device. In order to retrieve the optimal external displacements of the model, we apply a minimization strategy on the dissimilarity cost function.

### 3.3.5. Minimization Strategy

Regardless of the selected dissimilarity criteria, the cost function can always be expressed as follows:

$$C(\mathbf{q}) = (\mathbf{D}(I_t(\mathbf{M}\mathbf{q}_t) - I_{t_0}^*(\mathbf{M}\mathbf{q}_{t_0})))^2 \quad (18)$$

where  $\mathbf{q}_{t_i}$  denotes the vertex positions at time index  $t_i$ .  $\mathbf{D}$  represents the diagonal matrix that can be either the identity matrix  $\mathbf{I}$  or the weighting matrix  $\mathbf{H}_t$  depending on the selected criteria such that:

$$\mathbf{D} = \begin{cases} \mathbf{I}, & \text{if SSD, SCV} \\ \mathbf{H}_t, & \text{if WSSD, SCCV} \end{cases} \quad (19)$$

The term  $I_{t_0}^*$  represents the initial volume intensity and its value depends on the selected dissimilarity criteria. It can be computed as follows:

$$I_{t_0}^* = \begin{cases} I_{t_0}, & \text{if SSD, WSSD} \\ \widehat{I}_{t_0}^t, & \text{if SCV, SCCV} \end{cases} \quad (20)$$

If the SSD (or WSSD) criterion is used,  $I_{t_0}^*$  corresponds to the original initial image  $I_{t_0}$ . Otherwise,  $I_{t_0}^*$  represents the intensity of adapted reference volume  $\widehat{I}_{t_0}^t$  that allows matching intensity between current and initial image. It is worth mentioning that the computation of the adapted reference volume intensity  $\widehat{I}_{t_0}^t$  is expressed in Eq. 12. As demonstrated by Richa et al. (2011), the SCV and SSD criteria have the same behavior if there is no intensity variation. Furthermore, we can notice that all the dissimilarity criteria have similar performance when the quality of ultrasound images is good enough. This is because the confidence value of each voxel within the mesh model is higher than the minimum confidence threshold  $\tau$  ensuring  $\mathbf{H}_t = \mathbf{I}$ . Once the cost function is defined, the objective is to iteratively estimate the vertex displacements of the mesh model by minimizing the cost function  $C$ . To do so, we perform a Taylor expansion of the previous equation:

$$C(\mathbf{q}) \approx (\mathbf{D}\mathbf{J}\Delta\mathbf{q} + \mathbf{D}(I_t(\mathbf{M}\mathbf{q}_t^{k-1}) - I_{t_0}^*(\mathbf{M}\mathbf{q}_{t_0})))^2 \quad (21)$$

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 vertex 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:

$$\mathbf{J}(u, v) = \nabla I_t(u + wN_v) \mathbf{M}(u + w, v) \quad \text{with } w = v \bmod 3 \quad (22)$$

where  $u$  and  $v$  denote respectively a specific row and a specific column of the Jacobian matrix.  $\nabla I_t$  represent 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 gradient regarding the x, y and z axis. 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. Thus, the external displacements can be computed as follows:

$$\Delta \mathbf{q} = -\alpha \mathbf{J}^T \mathbf{D}^T \mathbf{D} [I_t(\mathbf{M}\mathbf{q}_t^{k-1}) - I_{t_0}^*(\mathbf{M}\mathbf{q}_{t_0})] \quad (23)$$

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

#### 4. Results

In the following sections, 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 experiments were conducted. The first type consists in testing our approach in nominal conditions where the target intensity is only affected by speckle noise and small intensity variation. Thus, we used the SSD criterion since we can assume that the US wave reflected by a physical point is constant and time independent. These first experiments allow demonstrating the relevance of the combination between mechanical model and intensity-based approach 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.

The second type of experiments allows validating our approach in intensity varying condition where the target is affected by large shadows and intensity shifts. For this purpose, we evaluated our method with different criteria on ultrasound images by using simulated data, and phantom data.

#### 4.1. 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. These steps can be executed offline in less than 3 minutes during an initialization phase. By using the method of Karamalis et al. (2012), the computation of the confidence image is performed in less than 40 ms. 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. We also set the confidence parameters such that  $\beta = 5$  and  $\tau = 85$ . 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 iteration 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.

#### 4.2. Validation in Nominal Conditions

In the following experiments, we demonstrate the performance of the original method that combines mechanical simulation and dense motion estimation on nominal conditions by using ultrasound images of good quality without shadows and ultrasound gain variation. Thus, we used the SSD criterion since all the presented similarity criteria have the same behavior as explained in section 3.3.5.

This 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 (Preiswerk et al., 2014; De Luca et al., 2015). 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 2 and 3. They are also presented in the challenge website<sup>1,2</sup>.

<b>Participants</b>	<b>Mean</b>	<b>SD</b>	<b>95%</b>
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 2: CLUST'14 results of 3D point-landmark tracking expressed in millimeters.

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

Table 3: 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 results for the different databases. The detailed results for all target tracking tasks are presented in table 4. As it can be seen in this 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 ( $<3$

<sup>1</sup>CLUST'14: <http://clust.ethz.ch/clust2014.html>

<sup>2</sup>CLUST'15: <http://clust.ethz.ch/results.html>

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 either ignoring the voxels positions that are outside the field of view in the cost function of Eq. (7) or by using prescan data as presented by (Royer et al., 2015). 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 observe that the error can be also 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 discussion section, we discuss the method robustness regarding the maximum amplitude of motion and deformation of the target.

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: 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. Validation with ultrasound gain changes

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.

#### 4.3.1. 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:

$$I_t(\mathbf{p}_k(t)) = I_t(\mathbf{p}_k(t)) + \kappa(t) \quad (24)$$

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.

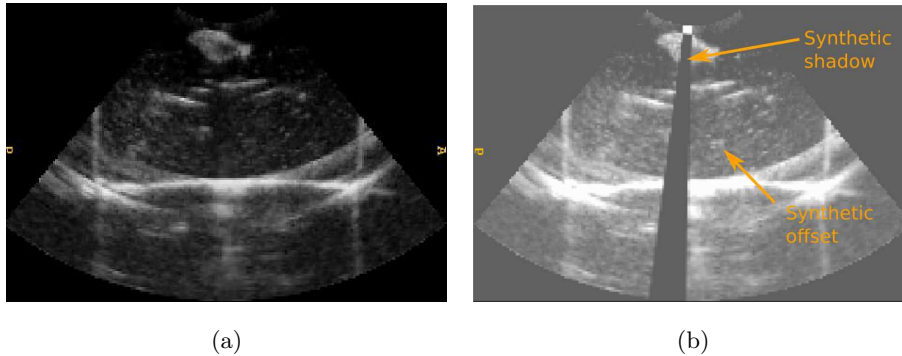


Figure 5: Simulation of US imaging gain change and shadows. (a) X-Y slice of original 3D ultrasound image. (b) X-Y slice of simulated 3D ultrasound image with shadows and gain change.

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.

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: 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. (Offset Min/Max) Minimum and maximum offsets applied to ultrasound 3D frames. (Nb Frames) Number of frame in the sequence

The key frames of these simulated ultrasound sequences are illustrated in Fig. 6 and Fig. 7. Fig. 6 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. 7 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.

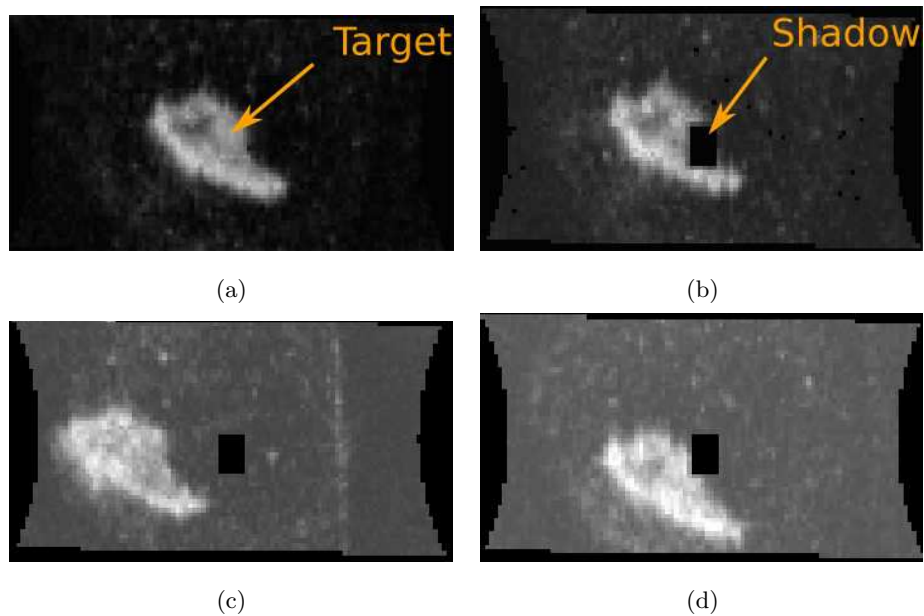


Figure 6: Illustration of sequence SIM\_3 at key frames. (a) Y-Z slice of frame 0 of sequence SIM\_1, SIM\_2, SIM\_3. (b) Y-Z slice of frame 6 of sequence SIM\_3. (c) Y-Z slice of frame 15 of sequence SIM\_3. (d) X-Z slice of frame 28 of sequence SIM\_3.



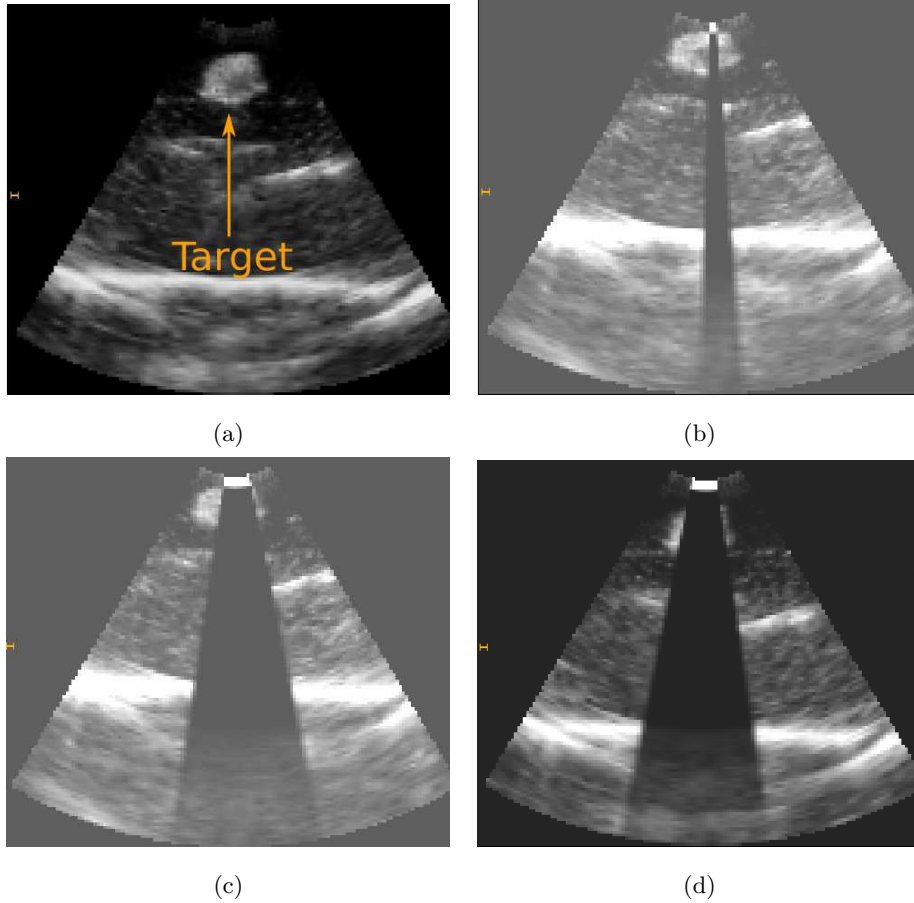


Figure 7: Illustration of sequences SIM\_4, SIM\_5, SIM\_6 at key frames. (a) X-Z slice of frame 0 of sequence SIM\_4, SIM\_5, SIM\_6. (b) X-Z slice of frame 25 of sequence SIM\_4. (c) X-Z slice of frame 25 of sequence SIM\_6. (d) X-Z slice of frame 42 of sequence SIM\_6.

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:

$$P_c = \frac{100}{N_v} \sum_{k=1}^{N_v} \mathbf{H}_t(\mathbf{p}_k(t)) \quad (25)$$

where  $N_v$  represents the total number of voxels in the mesh model.  $\mathbf{H}_t$  denotes the weight matrix expressed in Eq. (9). If  $P_c = 100\%$ , then all the voxels of the model are considered confident and not affected by shadows.

In Fig. 8, 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. 8, 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. 8, 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.

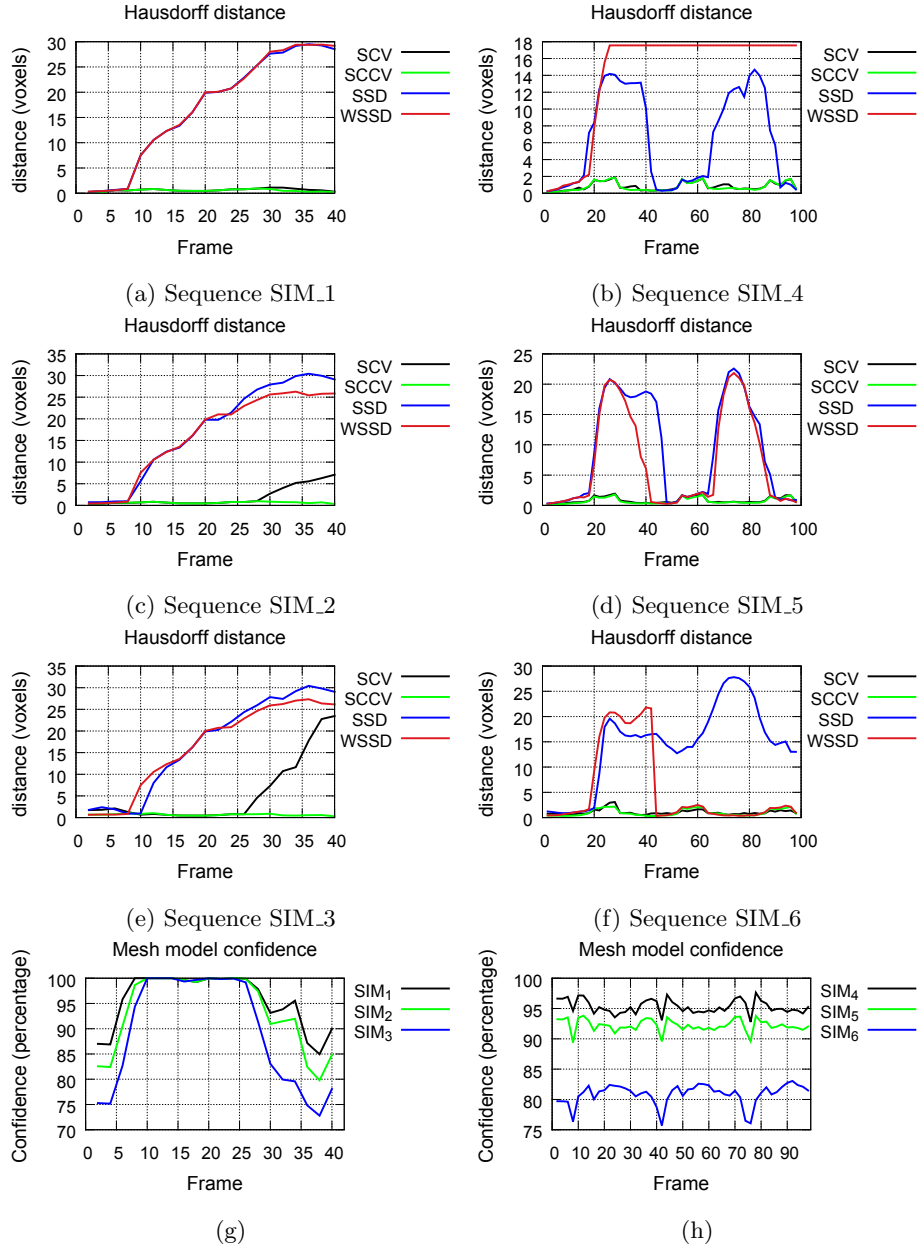


Figure 8: Evaluation of mean tracking error of criteria on each sequence of the simulated dataset.

In Figs. 9 and 10, 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. (12). From this latter, we observe 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. 9c and 9f 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. 9a) and current image (Fig. 9b). The intensity mapping function is therefore also straight and fits the ground truth curve. In Fig. 9f, 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. 9d) and current image (Fig. 9e). 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. 10c and 10f, 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. 10c, we show the high perturbation introduced by the shadow on JPDF computation from SCV criterion. In Fig. 10f, 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.

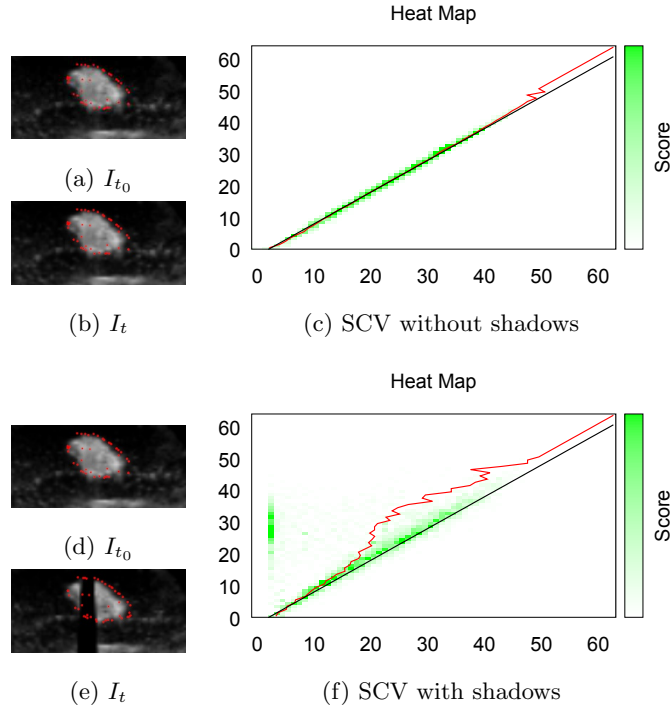


Figure 9: Illustration of synthetic shadow perturbation on the estimation of joint probability density function  $p_{I_t I_{t_0}}$  from SCV. (a-b) Reference and current images (d-e) Reference image and current image with synthetic shadow. (c-f) Comparison of the joint probability density function without and with shadows. The x-axis and y-axis represent respectively the intensity level of the target in the current image and in the reference image. The green cloud represents the joint probability density function. The red curve represents the intensity mapping function. The black curve denotes the ground truth shape of the intensity mapping function. Synthetic shadow perturbation can be observed in the left part ( $X = 4$  &  $20 < Y < 40$ ) of the figure (f).

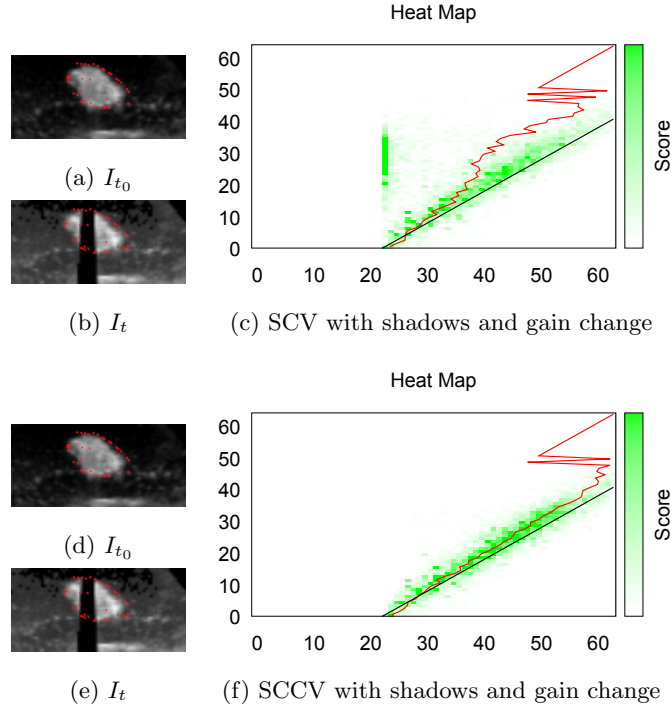


Figure 10: Illustration of synthetic shadow perturbation on the estimation of joint probability density function  $p_{I_t I_{t_0}}$  from SCV and SCCV. (a-d) Reference images. (b-e) current images with synthetic shadows. (c-f) Comparison of the joint probability density function from SCV and SCCV. The x-axis and y-axis represent respectively the intensity level of the target in the current image and in the reference image. The green cloud represents the joint probability density function. The red curve represents the intensity mapping function. The black curve denotes the ground truth shape of the intensity mapping function. Synthetic shadow perturbation can be observed in the middle part ( $X = 24$  &  $20 < Y < 40$ ) of the figure (c).

#### 4.3.2. Results on Phantom Data

We also evaluated the performance of 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 6. To provide ground truth, an expert annotated the position of one specific landmark in each volume of the sequences. The landmark corresponds to the extremum point of the target along its principal axis. In Figs. 11, 13 and 15, we show the initial landmark positions of the different sequences. From these annotations, we can estimate the mean tracking error by comparing the ground truth landmark positions and the warped point positions estimated from our tracking approach. The overall tracking results are shown in table 7.

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 6: Details of real 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	$9.4 \pm 7.3$	$4.7 \pm 4.0$	$19.1 \pm 18.1$	$2.3 \pm 1.5$
PHA_2	–	$6.6 \pm 4.8$	–	$2.4 \pm 0.9$
PHA_3	$40.2 \pm 16.0$	–	$1.4 \pm 0.8$	–
PHA_4	–	–	$12.8 \pm 19.1$	$1.5 \pm 0.6$

Table 7: 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.

To define the accuracy of annotations, we analyze the intra- and inter-observer variability of the landmarks for each sequence. To do so, three ob-

servers evaluated three times the landmarks positions along 10% of the total number of 3D frames. Then, the intra- and inter-variability are computed by measuring the mean error between the different sets of annotations. The results are shown in table 8. From table 7 and 8, we can deduce that the SCCV criterion provides accurate results since its mean tracking error remains approximately equivalent to the inter-variability. We do not measure the accuracy of PHA\_3 sequence because we assume that the landmark position remains constant since this sequence does not contain motion.

Sequence	Inter	Intra1	Intra2	Intra3
PHA_1	$2.3 \pm 0.9$	$1.3 \pm 0.7$	$1.4 \pm 0.8$	$1.2 \pm 0.6$
PHA_2	$2.5 \pm 1.0$	$1.7 \pm 0.9$	$1.7 \pm 1.0$	$1.5 \pm 0.7$
PHA_3	–	–	–	–
PHA_4	$1.1 \pm 0.7$	$0.7 \pm 0.4$	$0.8 \pm 0.4$	$0.9 \pm 0.4$

Table 8: Accuracy evaluation of manual annotation of landmark regarding each US sequence. The results are expressed in millimeters and represent respectively the mean error  $\pm$  the standard deviation (Sequence) Name of the sequence. (Inter) Inter-observer variability. (Intra1) Intra-observer variability for the first observer. (Intra2) Intra-observer variability for the second observer. (Intra3) Intra-observer variability for the third observer

The first sequence PHA\_1 allows comparing all the similarity criteria regarding the presence of large shadows and translation motion. Fig. 11 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 7, where we can observe that the mean tracking error remains small with the SCCV criterion. In Fig. 12, 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.



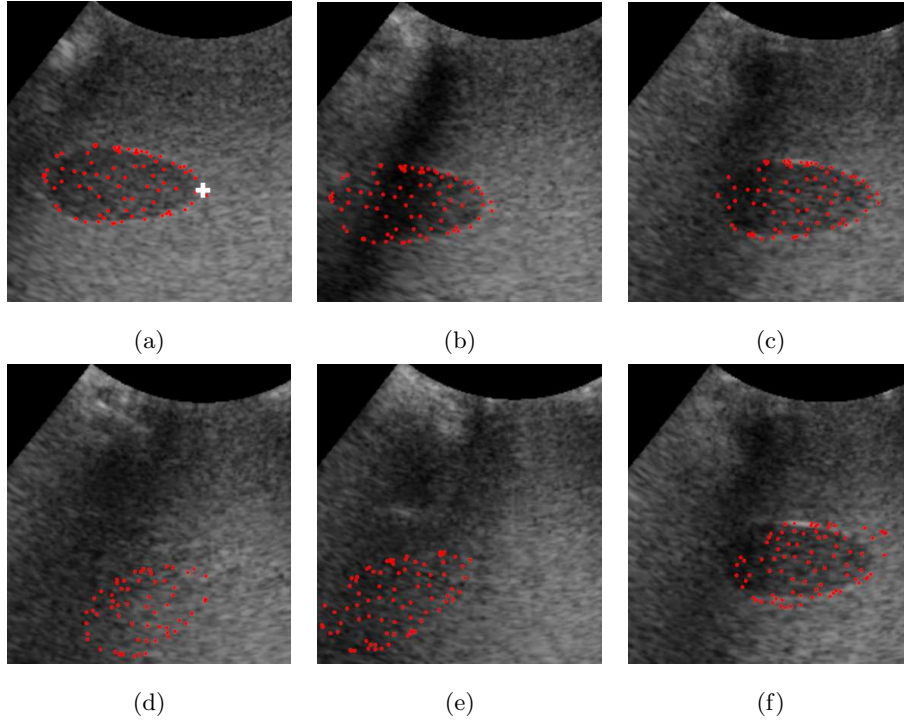


Figure 11: Example of tracking tasks during the PHA\_1 sequence with shadows and translation motion (see attached video). Red points represent the intersections between the mesh model and the X-Y slice of 3D image. The white cross represents the initial landmark selected by the expert. (a-b-c) SCCV tracking at time index 2 (a), at time index 52 (b), at time index 99 (c) of sequence PHA.2. (d) SSD tracking at time index 99, (e) SCV tracking at time index 99, (f) WSSD tracking at time index 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.

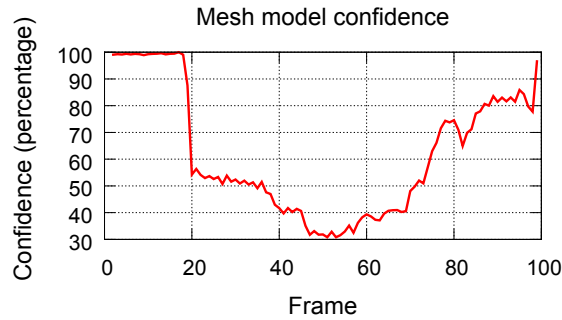


Figure 12: Evolution of the mesh model confidence during the sequence PHA.1. We can observe that the confidence of the mesh model decreases down to 30% at frame 50.

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. 13 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 7 that shows that the SCCV criterion achieves a smaller mean tracking error than the WSSD criterion.

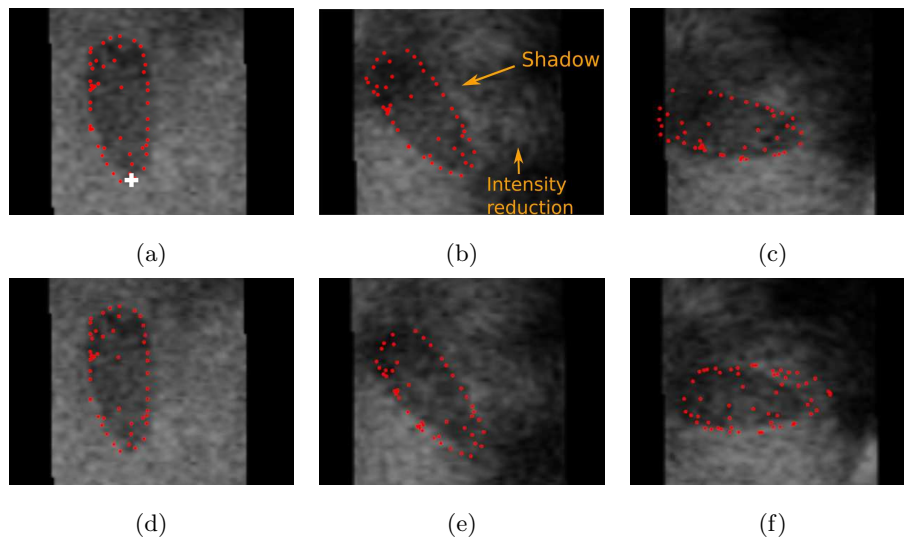


Figure 13: Example of tracking tasks during the PHA\_2 sequence with shadows and rotation motion. Red points represent the intersections between the mesh model and the Y-Z slice of 3D image (see attached video). The white cross represents the initial landmark selected by the expert. (a-b-c) SCCV tracking at frame 2 (a), at frame 122 (b), at frame 145 (c) of sequence PHA\_2. (d-e-f) WSSD tracking at frame 2 (d), at frame 122 (e), at frame 145 (f) 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 no annotation is required for this sequence since the target does not undergo any motion. The mean tracking error of this sequence is computed by comparing the current position of the mesh centroid with its initial position. 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. 14 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. Thus, we can observe that the SCV criterion achieves considerable better results than SSD from the table 7.

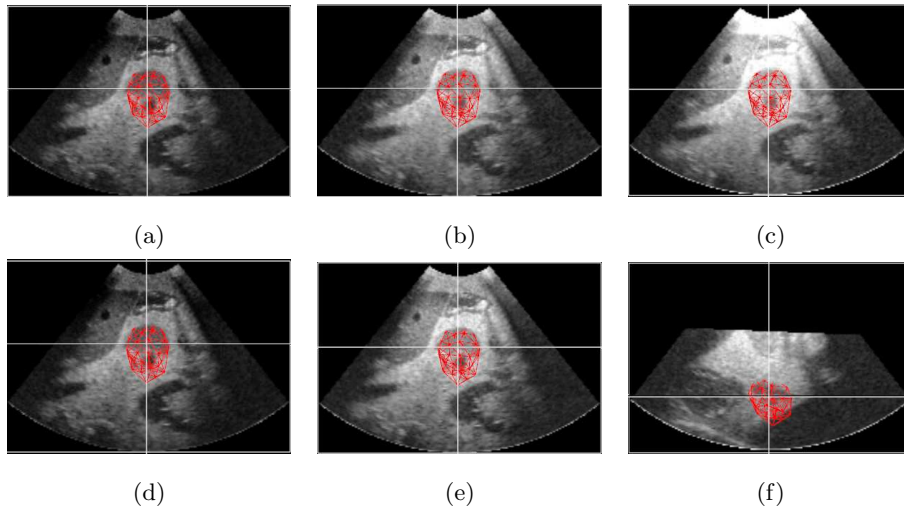


Figure 14: Example of tracking tasks during the PHA\_3 sequence with gain variation. Red model represents the 3D mesh model. (a-b-c) SCV tracking at frame 2 (a), at frame 8 (b), at frame 40 (c) of sequence PHA\_3. (d-e-f) SSD tracking at frame 2 (d), at frame 8 (e), at frame 40 (f) of sequence PHA\_3. 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. 15. 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 observe 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 7. Fig. 16 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.

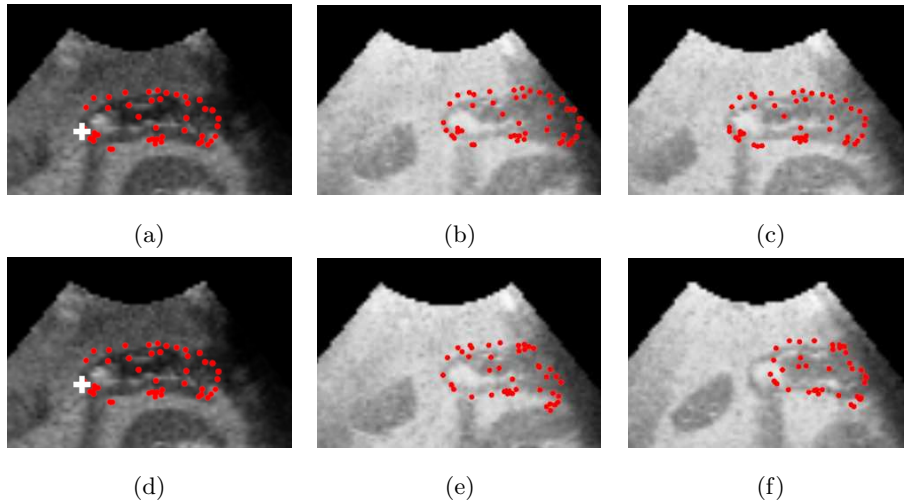


Figure 15: Example of tracking tasks during the PHA\_4 sequence with local and global intensity variations (see attached video). Red points represent the intersections between the mesh model and the image slice. The white crosses represent the initial landmarks selected by the expert. (a-b-c) SCCV tracking at frame 2 (a), at frame 51 (b), at frame 71 (c) of sequence PHA\_4. (d-e-f) SCV tracking at frame 2 (d), at frame 51 (e), at frame 71 (f) of sequence PHA\_4.

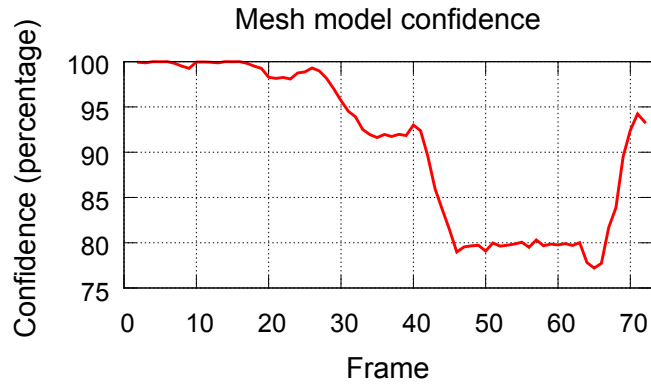


Figure 16: Evolution of the mesh model confidence during the sequence PHA\_4

From the previous experiments, we can deduce that the SSD and WSSD criteria are sensitive to imaging gain variation. To cope with that issue, we can use the SCV and SCCV criteria that are 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 13(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. 12).

## 5. Discussion

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 9. 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 9: 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:

$$M_{def} = \frac{d - D}{D} \quad (26)$$

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. 17, we illustrate the deformability measurements of different targets.

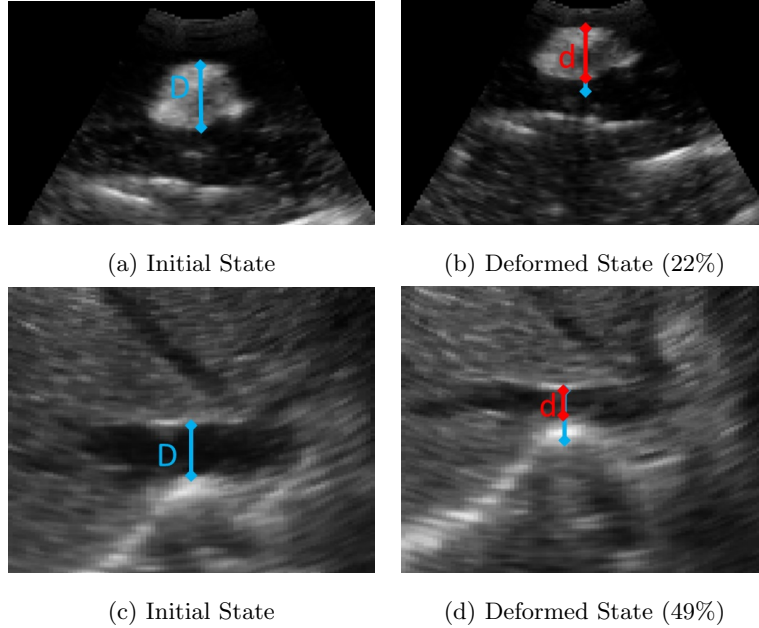


Figure 17: Example of deformable targets. (a-b) During the second experiment. (c-d) During the third experiment.

### 5.1. Sensitivity regarding $h_i$ parameter

Among the method's 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. 18, 19 and 20, 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. 18 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.

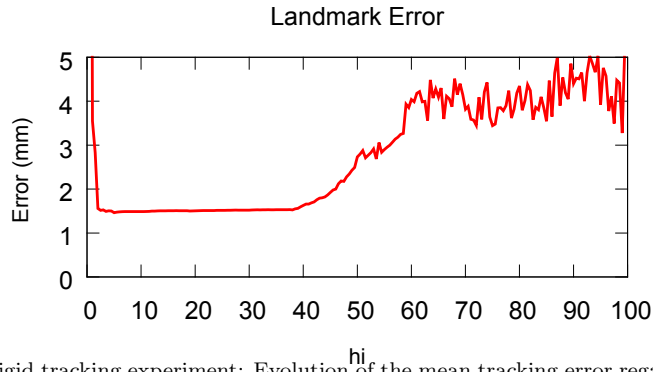


Figure 18: Rigid tracking experiment: Evolution of the mean tracking error regarding parameter  $h_i$

In Fig. 19, 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.

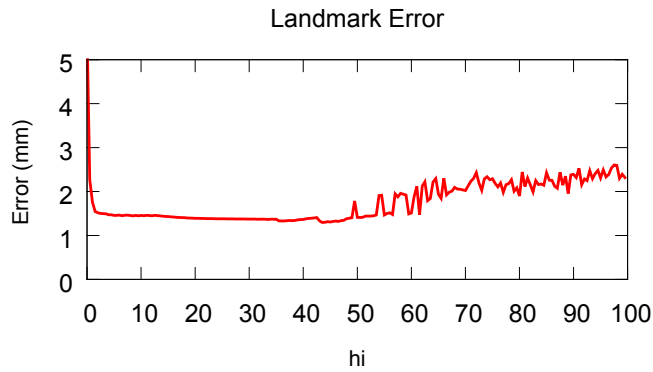


Figure 19: Tracking experiment with deformable (22%) target: Evolution of the mean tracking error regarding parameter



In the last experiment, we evaluate the sensitivity of  $h_i$  parameter on sequence that contains a highly deformable target (49%). Fig. 20 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.

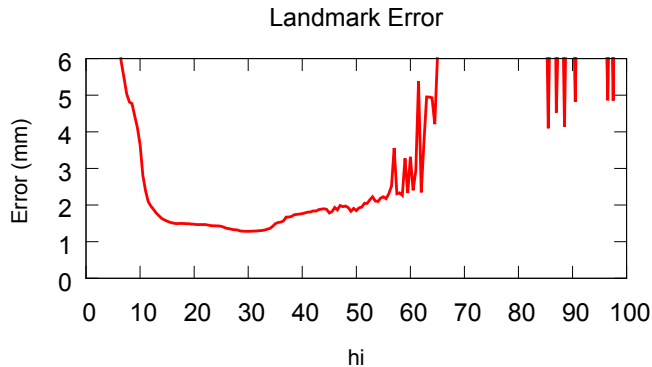


Figure 20: Tracking experiment with deformable (49%) target: Evolution of the mean tracking error regarding parameter

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.

### 5.2. Sensitivity regarding 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 9 (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. 21 shows respectively the MTE evolution regarding the Hausdorff distance on two sequences.

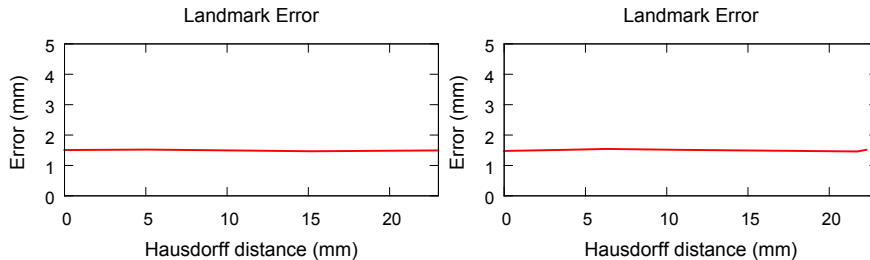


Figure 21: Evolution of the mean tracking error regarding the segmentation error (Left) Rigid target. (Right) Deformable target (22% of deformation)

As it can be seen from Fig. 21, 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.

### 5.3. Sensitivity regarding 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 9. In Fig. 22, we show the mean tracking error (y-axis) regarding the number of cells (x-axis) for rigid and non-rigid targets.

In Fig. 22, 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

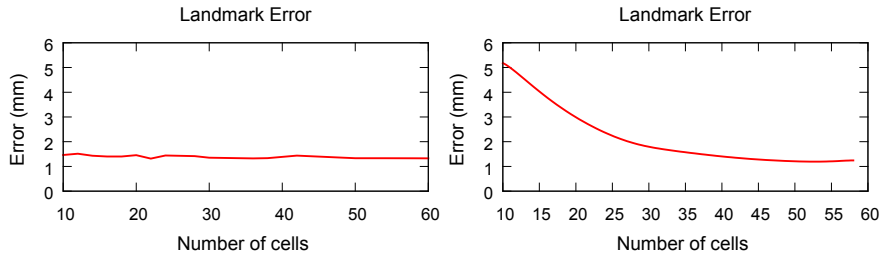


Figure 22: Evolution of the mean tracking error regarding the number of cells of the model (Left) Rigid target. (Right) Deformable target (22% of deformation)

same rigid displacement. In the non-rigid experiment showed in Fig. 22, 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.

#### 5.4. Sensitivity regarding 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 9 (PHA.5). In each experiment, the inter-volume target displacement is increased by removing consecutive US volumes from original sequence. In Fig. 23, 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.

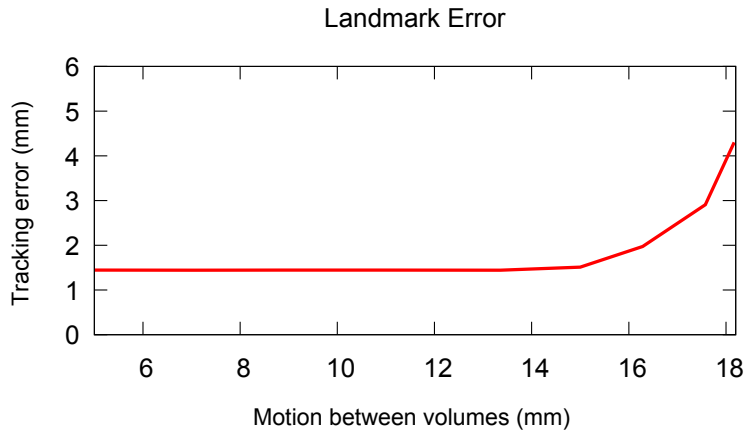


Figure 23: Rigid experiment: Tracking error regarding motion amplitude

### 5.5. Mechanical Regularization Benefit

Finally, we evaluate the benefit of mechanical regularization on the first sequence described in table 9. From this experiment, we compare the mean tracking error of the approach with and without mechanical regularization. The regularization term is removed by setting the parameters  $K_{ij}$ ,  $D_{ij}$  and  $G_i$  to zero. From that experiment, we observe that our method achieves a lower mean tracking error ( $1.5 \text{ mm} \pm 0.8 \text{ mm}$ ) compared to the method without mechanical contribution ( $4.7 \text{ mm} \pm 1.1 \text{ mm}$ ). Therefore, the mechanical regularization term allows providing better tracking performance as it is more robust to intensity variation introduced by noise. To see other results regarding the benefit of mechanical regularization, the interested reader is referred to our previous work (Royer et al. (2015)).

## 6. Conclusion

In this paper, we presented a 3D US target tracking approach that combines intensity-based method and mechanical simulation. We also proposed a new similarity criterion based on the quality measurement of US images. We first demonstrated that our approach provides high accuracy on real US images of the liver provided by MICCAI CLUST'14 and CLUST'15 challenges. Then, we

showed that this method provides good performance on US sequences that are perturbed by large shadows and ultrasound gain variation. The robustness and accuracy of the proposed method allow opening novel perspectives in procedures that involve deformable structures. For example, such approach can be well-suited for minimally-invasive therapies of the liver, e.g. radio-frequency ablation or high-intensity focused ultrasound. In these applications, the tumor position needs to be continuously tracked in order to adjust the needle position or ray beam angle over the time. Future work will consist in evaluating and adapting our method on tissues that undergo higher deformation such as cardiac muscle tissues.

## 7. Appendix

In this section, details are given about the computation of the optimal displacement parameters. Concerning the cost function expressed in Eq. 21, we have:

$$C(\Delta\mathbf{q}) \approx (\mathbf{D}\mathbf{J}\Delta\mathbf{q} + \mathbf{D}(I_t(\mathbf{M}\mathbf{q}_t^{k-1}) - I_{t_0}^*(\mathbf{M}\mathbf{q}_{t_0})))^2 \quad (27)$$

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:

$$\frac{\delta C(\Delta\mathbf{q})}{\delta\Delta\mathbf{q}} = 0 \quad (28)$$

It follows that:

$$\frac{\delta C(\Delta\mathbf{q})}{\delta\Delta\mathbf{q}} = 2(\mathbf{D}\mathbf{J})^T(\mathbf{D}\mathbf{J}\Delta\mathbf{q} + \mathbf{D}(I_t(\mathbf{M}\mathbf{q}_t^{k-1}) - I_{t_0}^*(\mathbf{M}\mathbf{q}_{t_0}))) = 0 \quad (29)$$

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

$$\Delta\mathbf{q} = (\mathbf{J}^T\mathbf{D}^T\mathbf{D}\mathbf{J})^{-1}\mathbf{J}^T\mathbf{D}^T\mathbf{D}(I_t(\mathbf{M}\mathbf{q}_t^{k-1}) - I_{t_0}^*(\mathbf{M}\mathbf{q}_{t_0})) \quad (30)$$

$$\Delta\mathbf{q} = -(\mathbf{D}\mathbf{J})^+\mathbf{D}(I_t(\mathbf{M}\mathbf{q}_t^{k-1}) - I_{t_0}^*(\mathbf{M}\mathbf{q}_{t_0})) \quad (31)$$

where  $(\mathbf{D}\mathbf{J})^+$  represents the left pseudo-inverse of the matrix  $\mathbf{D}\mathbf{J}$ . 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  $20000 \times 600$  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:

$$\mathbf{J}^T \mathbf{D}^T \mathbf{D} \mathbf{J} \approx \mathbf{J}^T \mathbf{D}^T \mathbf{D} \mathbf{J} + \lambda \mathbf{I} \quad (32)$$

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

$$\Delta \mathbf{q} = -\alpha \mathbf{J}^T \mathbf{D}^T \mathbf{D} [I_t(\mathbf{M}(\mathbf{q}^{k-1}(t))) - I_{t_0}^*(\mathbf{M}(\mathbf{q}(t_0)))] \quad (33)$$

### Acknowledgments

This research is currently supported by the Institute of Research and Technology (IRT) b-com.

### References

- Angelova, D., Mihaylova, L., 2010. Contour segmentation in 2d ultrasound medical images with particle filtering. *Machine Vision and Applications* 22, 551–561.
- Baker, S., Matthews, I., 2004. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision* 56, 221–255.
- Banerjee, J., Klink, C., Peters, E.D., Niessen, W.J., Moelker, A., van Walsum, T., 2015. Fast and robust 3d ultrasound registration Block and game theoretic matching. *Medical Image Analysis* 20, 173–183.
- Barbosa, D., Friboulet, D., Dhooge, J., 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* , 17–24.

- Basarab, A., Liebgott, H., Morestin, F., Lyshchik, A., Higashi, T., Asato, R., Delachartre, P., 2008. A method for vector displacement estimation with ultrasound imaging and its application for thyroid nodular disease. *Medical Image Analysis* 12, 259–274.
- Baumann, M., Mozer, P., Daanen, V., Troccaz, J., 2012. Prostate biopsy tracking with deformation estimation. *Medical image analysis* 16, 562–576.
- Bell, M.A.L., Byram, B.C., Harris, E.J., Evans, P.M., 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, 1359–1374.
- Berge, C.S.z., Kapoor, A., Navab, N., 2014. Orientation-Driven Ultrasound Compounding Using Uncertainty Information, in: *Proc. of International Conference on Information Processing in Computer-Assisted Interventions*, Springer International Publishing, Cham. pp. 236–245.
- Chang, R.F., Wu, W.J., Moon, W.K., Chen, D.R., 2005. Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors. *Breast Cancer Research and Treatment* 89, 179–185.
- Chatelain, P., Krupa, A., Navab, N., 2015. Optimization of ultrasound image quality via visual servoing, in: *Proc. of IEEE International Conference on Robotics and Automation*, pp. 5997–6002.
- Cohen, B., Dinstein, I.h., 2002. New maximum likelihood motion estimation schemes for noisy ultrasound images. *Pattern Recognition* 35, 455–463.
- De Luca, V., Benz, T., Kondo, S., Knig, L., Lbke, D., Rothlbbbers, S., Somphone, O., Allaire, S., Lediju Bell, M.A., Chung, D.Y.F., Cifor, A., Grozea, C., Gnther, M., Jenne, J., Kipshagen, T., Kowarschik, M., Navab, N., Rhaak, J., Schwaab, J., Tanner, C., 2015. The 2014 liver ultrasound tracking benchmark. *Physics in Medicine and Biology* 60, 5571–5599.
- De Luca, V., Tschannen, M., Szkely, G., 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, Springer. pp. 518–525.
- Elen, A., Hon Fai Choi, Loeckx, D., Hang Gao, Claus, P., Suetens, P., Maes, F., 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, 1580–1591.
- Hellier, P., Coup, P., Morandi, X., Collins, D.L., 2010. An automatic geometrical and statistical method to detect acoustic shadows in intraoperative ultrasound brain images. *Medical Image Analysis* 14, 195–204.
- Heyde, B., Claus, P., Jasaityte, R., Barbosa, D., Bouchez, S., Vandenheuevel, M., Wouters, P., Maes, F., Hooge, J.D., 2012. Motion and deformation estimation of cardiac ultrasound sequences using an anatomical B-spline transformation model, in: Proc. of IEEE International Symp. on Biomedical Imaging, pp. 266–269.
- Higgins, H., Berger, D.L., 2006. RFA for Liver Tumors: Does It Really Work? *The Oncologist* 11, 801–808.
- Karamalis, A., Wein, W., Klein, T., Navab, N., 2012. Ultrasound confidence maps using random walks. *Medical Image Analysis* 16, 1101–1112.
- Ledesma-Carbayo, M.J., Kybic, J., Desco, M., Santos, A., 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, Springer. pp. 889–896.
- Lediju, M., Byram, B.C., Harris, E.J., Evans, P.M., Bamber, J.C., others, 2010. 3d Liver tracking using a matrix array: Implications for ultrasonic guidance of IMRT, in: Proc. of IEEE Ultrasonics Symp., pp. 1628–1631.
- Lee, D., 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, pp. 2831–2836.
- Loosvelt, M., Villard, P.F., Berger, M.O., 2014. Using a biomechanical model for tongue tracking in ultrasound images, in: Proc. of IEEE Symp. on Biomedical Simulation, Springer. pp. 67–75.
- Lubke, D., Grozea, C., 2014. High Performance Online Motion Tracking in Abdominal Ultrasound Imaging, in: Proc. of MICCAI Workshop on Challenge on Liver Ultrasound Tracking, p. 29.
- Marami, B., Sirouspour, S., Fenster, A., W. Capson, D., 2014. Dynamic Tracking of a Deformable Tissue Based on 3d-2d MR-US Image Registration, in: Proc. of SPIE Medical Imaging.
- Masum, M.A., Pickering, M., Lambert, A., Scarvell, J., Smith, P., 2014. Accuracy assessment of Tri-plane B-mode ultrasound for non-invasive 3d kinematic analysis of knee joints. *Biomedical engineering online* 13, 122.
- Metz, C., Klein, S., Schaap, M., van Walsum, T., Niessen, W., 2011. Nonrigid registration of dynamic medical imaging data using nD+t B-splines and a groupwise optimization approach. *Medical Image Analysis* 15, 238–249.
- Mikic, I., Krucinski, S., Thomas, J.D., 1998. Segmentation and tracking in echocardiographic sequences: active contours guided by optical flow estimates. *IEEE Trans. on Medical Imaging* 17, 274–284.
- Mukherjee, R., Sprouse, C., Abraham, T., Hoffmann, B., McVeigh, E., Yuh, D., Burlina, P., 2011. Myocardial motion computation in 4d ultrasound, in: Proc. of IEEE International Symp. on Biomedical Imaging, IEEE. pp. 1070–1073.
- Nickolls, J., Buck, I., Garland, M., Skadron, K., 2008. Scalable parallel programming with CUDA 6, 40–53.
- Papademetris, X., Sinusas, A., Dione, D., Constable, R., Duncan, J., 2002. Estimation of 3-D left ventricular deformation from medical images using biomechanical models. *IEEE Transactions on Medical Imaging* 21, 786–800.

- Penneç, X., Cachier, P., Ayache, N., 2001. Tracking brain deformations in time-sequences of 3d US images, in: Proc. of International Conference on Information Processing in Medical Imaging, Springer. pp. 169–175.
- Penney, G., Blackall, J., Hamady, M., Sabharwal, T., Adam, A., Hawkes, D., 2004. Registration of freehand 3d ultrasound and magnetic resonance liver images. *Medical Image Analysis* 8, 81–91, doi:10.1016/j.media.2003.07.003.
- Pernot, M., Tanter, M., Fink, M., 2004. 3-D real-time motion correction in high-intensity focused ultrasound therapy. *Ultrasound in Medicine & Biology* 30, 1239–1249.
- Pickering, M.R., Muhiç, A., Scarvell, J.M., Smith, P.N., 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, pp. 5821–5824.
- Preiswerk, F., De Luca, V., Arnold, P., Celicanin, Z., Petrusca, L., Tanner, C., Bieri, O., Salomir, R., Cattin, P.C., 2014. Model-guided respiratory organ motion prediction of the liver from 2d ultrasound. *Medical Image Analysis* 18, 740–751.
- Richa, R., Souza, M., Scandaroli, G., Comunello, E., 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, pp. 373–377.
- Richa, R., Sznitman, R., Taylor, R., Hager, G., 2011. Visual tracking using the sum of conditional variance, in: Proc. of IEEE International Conference on Intelligent Robots and Systems, pp. 2953–2958.
- Rothlubbers, S., Schwaab, J., Jenne, J., Gunther, M., 2014. MICCAI CLUST 2014-Bayesian Real-Time Liver Feature Ultrasound Tracking, in: Proc. of MICCAI Workshop on Challenge on Liver Ultrasound Tracking, p. 45.

- Royer, L., Marchal, M., Le Bras, A., Dardenne, G., Krupa, A., 2015. Real-time Tracking of Deformable Target in 3d Ultrasound Images, in: Proc. of IEEE International Conference on Robotics and Automation.
- Schneider, R.J., Perrin, D.P., Vasilyev, N.V., Marx, G.R., del Nido, P.J., Howe, R.D., 2012. Real-time image-based rigid registration of three-dimensional ultrasound. *Medical Image Analysis* 16, 402–414.
- Schroeder, W., Martin, K.W., Lorensen, B., 2002. The visualization toolkit: an object-oriented approach to 3D graphics. Kitware, Inc.
- Shekhar, R., Zagrodsky, V., 2002. Mutual information-based rigid and nonrigid registration of ultrasound volumes. *IEEE Transactions on Medical Imaging* 21, 9–22.
- Si, H., 2015. TetGen, a Delaunay-Based Quality Tetrahedral Mesh Generator. *ACM Trans. on Mathematical Software* 41.
- Somphone, O., Allaire, S., Mory, B., Dufour, C., 2014. Live Feature Tracking in Ultrasound Liver Sequences with Sparse Demons, in: Proc. of MICCAI Workshop on Challenge on Liver Ultrasound Tracking, p. 53.
- Touil, B., Basarab, A., Delachartre, P., Bernard, O., Friboulet, D., 2010. Analysis of motion tracking in echocardiographic image sequences: Influence of system geometry and point-spread function. *Ultrasonics* 50, 373–386.
- Veronesi, F., Corsi, C., Caiani, E.G., Lamberti, C., 2005. Nearly automated left ventricular long axis tracking on real time three-dimensional echocardiographic data, in: IEEE International Conference on Computers in Cardiology, pp. 5–8.
- Vijayan, S., Klein, S., Hofstad, E.F., Lindseth, F., Ystgaard, B., Lango, T., 2013. Validation of a non-rigid registration method for motion compensation in 4d ultrasound of the liver, in: Proc. of IEEE International Symp. on Biomedical Imaging, pp. 792–795.

- Wein, W., Cheng, J.Z., Khamene, A., 2008. Ultrasound based respiratory motion compensation in the abdomen, in: Proc. of MICCAI Workshop on Image Guidance and Computer Assistance for Soft tissue Interventions, p. 294.
- Yeung, F., Levinson, S.F., Fu, D., Parker, K.J., 1998. Feature-adaptive motion tracking of ultrasound image sequences using a deformable mesh. *IEEE Transactions on Medical Imaging* 17, 945–956.
- Yipeng Hu, Carter, T.J., Ahmed, H.U., Emberton, M., Allen, C., Hawkes, D.J., Barratt, D.C., 2011. Modelling Prostate Motion for Data Fusion During Image-Guided Interventions. *IEEE Transactions on Medical Imaging* 30, 1887–1900.
- Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3d active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage* 31, 1116–1128.
- Zhang, X., Gnther, M., 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, Springer. pp. 286–294.