

Intensity-based visual servoing for non-rigid motion compensation of soft tissue structures due to physiological motion using 4D ultrasound

Deukhee Lee and Alexandre Krupa

Abstract—This paper presents a visual-servoing method for compensating motion of soft tissue structures using 4D ultrasound. The motion of soft tissue structures caused by physiological and external motion makes it difficult to investigate them for diagnostic and therapeutic purposes. The main goal is to track non-rigidly moving soft tissue structures and compensate the motion in order to keep a lesion on its target position during a treatment. We define a 3D non-rigid motion model by extending the Thin-Plate Spline (TPS) algorithm. The motion parameters are estimated with intensity-value changes of a points set in a tracking soft tissue structure. Finally, the global rigid motion is compensated with a 6-DOF robot according to the motion parameters of the tracking structure. Simulation experiments are performed with recorded 3D US images of *in-vivo* soft tissue structures and validate the effectiveness of the non-rigid motion tracking method. Robotic experiments demonstrated the success of our method with a deformable phantom.

I. INTRODUCTION

Medical imaging modalities make it possible to observe soft tissue structures non-invasively. Among them, ultrasound (US) imaging has many benefits. It is cheap, real-time, safe for human body, and non-interactive with ferromagnetic materials. For these reasons, US is the most widespread medical imaging modality.

Physiological motions such as respiration and heartbeat move soft tissue structures globally and deform them locally. Therefore this motion of soft tissue structures makes it difficult to investigate them for diagnostic and therapeutic purposes. Especially the target position of non-invasive therapy should follow the physiological motion of a moving lesion.

There are some literatures to deal with motion compensation using US imaging. In [1] and [2], speckle decorrelation is used to estimate elevational motion of 2D US probe, and Krupa et al.[2] successfully compensate the soft tissue rigid motion with a 2D probe attached to a robot. In [3], an US image-based visual servoing method is presented to position a 2D probe on a desired organ section and track it thanks to the use of image moments obtained from contour segmentation. Nadeau tracks 3D rigid motion using a virtual 3D US probe in [4]. In [5], 3D translational motions are estimated using 4D US. However, non-rigid motion is not considered in the above methods.

The contributions of this paper are to track 3D non-rigid motion using 4D US in real time and to compensate the motion. At our best knowledge, unlike other non-rigid

registration procedures, our method is processed in real time. In the rest of this paper, a non-rigid motion model is defined using 3D Thin-Plate Spline (TPS). Then, an intensity-based tracking method is described to estimate the parameters of the motion model. Global rigid motion such as 3 translations and 3 rotations are extracted from non-rigid motion parameters. Finally, a 6 degree-of-freedom (DOF) robot equipped with a 4D US probe is controlled to compensate the rigid motion. Tracking accuracy is discussed from simulation experiments using *in-vivo* US images and the effectiveness of the proposed method is also verified from robotic experiments with a deformable tissue-mimicking (TM) phantom.

II. NON-RIGID MOTION TRACKING

Before compensating the 3D motion of soft tissue structures, a new method to estimate 3D motion of deformable soft tissue is proposed in this section.

A. 3D Scan Conversion

We use a *SonixTOUCH Research* 4D US scanner (Ultrasound Medical Corporation, Canada) and a motorized US 3D transducer (Model: 4DC7-3/40, Ultrasonix Medical Corporation, Canada). Since the 4D US scanner is designed for the purpose of research works, we can access digital data before the conversion into an image. Afterwards, a set of volume data is converted into a volumetric image, called as 3D scan conversion.

For scan conversion, geometries of the probe (the 2D US transducer's radius R_{probe} and the motor's radius R_{motor}) and imaging parameters (the number of sample data in a A-line $N_{samples}$, the number of A-lines in a frame N_{lines} , and the number of frames in a scan volume N_{frames} , angle between neighboring A-lines α_{line} , and angle between neighboring frames α_{frame}) are considered in (1). In Fig. 1, a sample $s(i, j, k)$, which is the i -th datum along the j -th A-line in the k -th frame, is relocated into a point $\mathbf{p}(r, \varphi, \theta)$ in a volumetric image, which is represented as $\mathbf{p}(x, y, z)$ in Cartesian coordinates according to (1) and (2).

Note that our 3D US probe continuously scans volumes while its motor is sweeping the volume in forward and backward directions. Additionally, sweeping direction d (which is 1 in the forward direction and 0 in the backward direction in (1c)) and motor's rotating motion should be considered. In (1c), we assumed that the motor stops during the time to scan A-line so that scan lines are straight.

The authors are with INRIA Rennes-Bretagne Atlantique, IRISA, Campus de Beaulieu, 35042 Rennes cedex, France Deukhee.Lee, Alexandre.Krupa@inria.fr

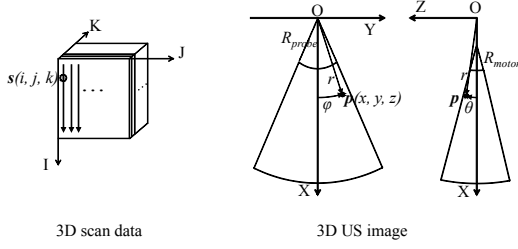


Fig. 1. 3D scan data (left) are reconstructed into a 3D ultrasound image (right).

$$r = D_{sample} \times i + R_{probe} \quad (1a)$$

$$\varphi = -0.5\alpha_{line}(N_{lines} - 1) + \alpha_{line}j \quad (1b)$$

$$\theta = (-1)^d \left\{ \frac{\alpha_{frame}N_{frames}}{2} - \frac{\alpha_{frame}N_{frames}}{N_{lines}N_{frames} - 1} (j + N_{lines}k) \right\} \quad (1c)$$

In (1a), D_{sample} means the distance between neighboring two sample data.

$$x = (r \cos \varphi - C) \cos \theta + C \quad (2a)$$

$$y = r \sin \varphi \quad (2b)$$

$$z = (r \cos \varphi - C) \sin \theta \quad (2c)$$

In (2), C is the offset distance from the probe's origin to the motor's origin, $C = R_{probe} - R_{motor}$.

B. Non-Rigid Motion Model

Soft tissue structures move and deform due to physiological motion. Therefore, we can model the motion of soft tissue as an affine motion part and a deformation part as shown in (3):

$$\begin{bmatrix} \mathbf{p}' \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix} + \begin{bmatrix} \mathbf{n} \\ 0 \end{bmatrix} \quad (3)$$

where a point $\mathbf{p} = (x, y, z)^T$ moves to a point $\mathbf{p}' = (x', y', z')^T$ according to the affine motion matrix (the first matrix in the right side of the equation) and the deformation vector $\mathbf{n} = (n_x, n_y, n_z)^T$.

The deformation vector \mathbf{n} is modeled using TPS ([6]):

$$\mathbf{n} = \sum_{i=1}^n \mathbf{w}_i U(\mathbf{c}_i - \mathbf{p}) \quad (4)$$

where \mathbf{n} is defined with a number of n control points $\mathbf{c}_i = (\check{x}_i, \check{y}_i, \check{z}_i)^T$ and their weights $\mathbf{w}_i = (w_i^x, w_i^y, w_i^z)^T$. The base function $U(s)$ usually denotes the Euclidean norm of s in 3D cases.

As a result, the non-rigid motion model (3) is the three-dimensional extended version ($\mathfrak{R}^3 \rightarrow \mathfrak{R}^3$) of TPS warping models ($\mathfrak{R}^2 \rightarrow \mathfrak{R}^2$ and $\mathfrak{R}^2 \rightarrow \mathfrak{R}^3$, respectively) presented in [7] and [8]. The parameters of the motion model, (\mathbf{A} , \mathbf{t} , and $\mathbf{W} = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$), are estimated with a set of points $\mathbf{P} = \{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ in the reference image and the corresponding set of points $\mathbf{P}' = \{\mathbf{p}'_1, \dots, \mathbf{p}'_n\}$ in sequential images. As the

set of initial control points \mathbf{C} moves to the corresponding set of control points \mathbf{C}' , the points $\mathbf{p} \in \mathbf{C}$ also move to the points $\mathbf{p}' \in \mathbf{C}'$. Therefore, the parameters are given by the relationship:

$$\begin{bmatrix} \mathbf{L} & \mathbf{C} \\ \mathbf{C}^T & \mathbf{0}_{4 \times 4} \end{bmatrix} [\mathbf{W} \quad \mathbf{t} \quad \mathbf{A}]^T = \begin{bmatrix} \mathbf{C}' \\ \mathbf{0}_{4 \times 3} \end{bmatrix} \quad (5)$$

where \mathbf{L} is a $n \times n$ matrix, whose element is $L_{ij} = U(\mathbf{c}_j - \mathbf{c}_i)$.

$\mathbf{C} = \begin{bmatrix} 1 & \check{x}_1 & \check{y}_1 & \check{z}_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \check{x}_n & \check{y}_n & \check{z}_n \end{bmatrix}$ is a initial $n \times 4$ control points

matrix and $\mathbf{C}' = [\mathbf{c}'_1 \quad \dots \quad \mathbf{c}'_n]^T$ is a $n \times 3$ destination control points matrix.

According to [6], the motion model (3) is rewritten as (6).

$$\mathbf{p}'^T = \mathbf{M}(\mathbf{p}) \mathbf{K}_*(\mathbf{C}) \mathbf{C}' \quad (6)$$

where $\mathbf{M}(\mathbf{p}) = [U(\mathbf{p} - \mathbf{c}_1) \quad \dots \quad U(\mathbf{p} - \mathbf{c}_n) \quad 1 \quad x \quad y \quad z]$ and $\mathbf{K}_*(\mathbf{C})$ denotes the $(n+4) \times n$ sub-matrix of the inverted form of the leftmost matrix in (5).

When a set of control points \mathbf{C} and a set of point \mathbf{P} are initialized as \mathbf{C}_0 and \mathbf{P}_0 respectively, $\mathbf{M}(\mathbf{P}_0)$ and $\mathbf{K}_*(\mathbf{C}_0)$ in (6) become constant matrices. Therefore, the corresponding points \mathbf{P}' obtained after deformation are just up to the set of moving control points \mathbf{C}' . This means that the control points can be considered as the motion parameters of the non-rigid motion model:

$$\mathbf{P}^T = \mathbf{M}(\mathbf{P}_0) \mathbf{K}_*(\mathbf{C}_0) \mathbf{C}$$

C. Intensity-Based Tracking

Now, we propose to use the parametric motion model defined in the previous section to track a deformable object from a sequence of 3D US images. Based on the same principle of the 2D region tracking method presented in [9] we define a parametric motion model:

$$\mathbf{p} = f(\mathbf{p}_0, \boldsymbol{\mu}(t)) \quad (7)$$

where \mathbf{p}_0 and \mathbf{p} represent respectively the initial and transformed coordinates of a 3D point, and $\boldsymbol{\mu}(t)$ is a motion parameters vector at time t . In our non-rigid motion model (6), we define $\boldsymbol{\mu}(t) = [\check{x}_1(t) \quad \check{y}_1(t) \quad \check{z}_1(t) \quad \dots \quad \check{x}_n(t) \quad \check{y}_n(t) \quad \check{z}_n(t)]^T$ from a set of n control points \mathbf{C} at time t .

We define a tracking region with a set of intensity values at several 3D points \mathbf{P} within the target region in a volumetric image. This tracking region \mathbf{I} is a function of a set of initial coordinates of \mathbf{P} and the $\boldsymbol{\mu}$ vector at time t given by:

$$\mathbf{I}(\mathbf{P}_0, \boldsymbol{\mu}(t)) = \begin{bmatrix} I(f(\mathbf{p}_1^0, \boldsymbol{\mu}(t))) \\ I(f(\mathbf{p}_2^0, \boldsymbol{\mu}(t))) \\ \vdots \\ I(f(\mathbf{p}_N^0, \boldsymbol{\mu}(t))) \end{bmatrix} \quad (8)$$

where $I(\mathbf{p}_i)$ denotes the intensity value at the location $\mathbf{p}_i = (x_i, y_i, z_i)^T$ at time t and N is the number of 3D points.

Now, we define a Jacobian matrix \mathbf{J}_μ , which links the time variation of the tracking region \mathbf{I} to the time variation of the motion parameters vector μ :

$$\frac{\partial \mathbf{I}}{\partial t} = \mathbf{J}_\mu \frac{d\mu}{dt} = \mathbf{J}_\mu \mathbf{v}_\mu \quad (9)$$

where \mathbf{v}_μ is the velocity vector of motion parameters.

The Jacobian matrix \mathbf{J}_μ can be calculated from the motion model, $\mathbf{p} = f(\mathbf{p}_0, \mu(t))$, and the gradient values $\nabla_{\mathbf{p}_0} \mathbf{I}$ of the tracking region \mathbf{I} at time t as shown in (10).

$$\mathbf{J}_\mu = \frac{\partial \mathbf{I}}{\partial \mu} = \frac{\partial \mathbf{I}}{\partial \mathbf{P}_0} \times \frac{\partial \mathbf{P}_0}{\partial \mathbf{P}} \times \frac{\partial \mathbf{P}}{\partial \mu} = \nabla_{\mathbf{P}_0} \mathbf{I} \times f_{\mathbf{P}_0}^{-1} \times f_\mu \quad (10)$$

where the gradient values $\nabla_{\mathbf{p}_0} \mathbf{I}$ can be measured using a $3 \times 3 \times 3$ or $5 \times 5 \times 5$ Sobel filter. $f_{\mathbf{P}_0}$ and f_μ are calculated according to (11) and (12).

$$f_{\mathbf{p}_0} = \begin{bmatrix} \frac{\partial x}{\partial x_0} & \frac{\partial x}{\partial y_0} & \frac{\partial x}{\partial z_0} \\ \frac{\partial y}{\partial x_0} & \frac{\partial y}{\partial y_0} & \frac{\partial y}{\partial z_0} \\ \frac{\partial z}{\partial x_0} & \frac{\partial z}{\partial y_0} & \frac{\partial z}{\partial z_0} \end{bmatrix} = \left(\begin{bmatrix} \frac{\partial \mathbf{M}}{\partial x_0} \\ \frac{\partial \mathbf{M}}{\partial y_0} \\ \frac{\partial \mathbf{M}}{\partial z_0} \end{bmatrix} \mathbf{K}_* \mathbf{C} \right)^T \quad (11)$$

$$f_\mu = \begin{bmatrix} \frac{\partial x}{\partial \tilde{x}_1} & \frac{\partial x}{\partial \tilde{y}_1} & \frac{\partial x}{\partial \tilde{z}_1} & \dots & \frac{\partial x}{\partial \tilde{x}_n} & \frac{\partial x}{\partial \tilde{y}_n} & \frac{\partial x}{\partial \tilde{z}_n} \\ \frac{\partial y}{\partial \tilde{x}_1} & \frac{\partial y}{\partial \tilde{y}_1} & \frac{\partial y}{\partial \tilde{z}_1} & \dots & \frac{\partial y}{\partial \tilde{x}_n} & \frac{\partial y}{\partial \tilde{y}_n} & \frac{\partial y}{\partial \tilde{z}_n} \\ \frac{\partial z}{\partial \tilde{x}_1} & \frac{\partial z}{\partial \tilde{y}_1} & \frac{\partial z}{\partial \tilde{z}_1} & \dots & \frac{\partial z}{\partial \tilde{x}_n} & \frac{\partial z}{\partial \tilde{y}_n} & \frac{\partial z}{\partial \tilde{z}_n} \end{bmatrix} \quad (12)$$

$$= \begin{bmatrix} c_1 & 0 & 0 & \dots & c_n & 0 & 0 \\ 0 & c_1 & 0 & \dots & 0 & c_n & 0 \\ 0 & 0 & c_1 & \dots & 0 & 0 & c_n \end{bmatrix}$$

where c_i is the i -th element of $(\mathbf{M}\mathbf{K}_*)^T$.

The strategy of the intensity-based region tracking is to minimize the error $\mathbf{e}(\mathbf{P}_0, \mu(t)) = \mathbf{I}(\mathbf{P}_0, \mu(t)) - \mathbf{I}^*$, where \mathbf{I}^* is the reference target region fixed at time t_0 : $\mathbf{I}^* = \mathbf{I}(\mathbf{P}_0, \mu(t_0))$.

Considering an exponential decrease of the error, let the time variation of the error $\dot{\mathbf{e}} = -\lambda \mathbf{e}$ where λ is the proportional coefficient involved in the exponential convergence of \mathbf{e} . $\dot{\mathbf{e}}$ is the same with the time variation of the tracking region: $\dot{\mathbf{e}} = \frac{\partial \mathbf{I}}{\partial t}$. As a result, the velocity vector of motion parameters is given by:

$$\mathbf{v}_\mu = -\lambda \mathbf{J}_\mu^+ (\mathbf{I} - \mathbf{I}^*) \quad (13)$$

where \mathbf{J}_μ^+ is the Moore-Penrose pseudo-inverse of \mathbf{J}_μ , that is $\mathbf{J}_\mu^+ = (\mathbf{J}_\mu^T \mathbf{J}_\mu)^{-1} \mathbf{J}_\mu^T$ if $N > n$. Finally, motion parameters are estimated as $\mu(t + \tau) = \mu(t) + \mathbf{v}_\mu \tau$ where τ is the sampling period of tracking processes.

$\nabla_{\mathbf{P}_0} \mathbf{I}$ and $f_{\mathbf{P}_0}$ changes according to $\mu(t)$ even though \mathbf{P}_0 is initialized and fixed. In practice, it is time-consuming to calculate \mathbf{J}_μ and \mathbf{J}_μ^+ every time because the size $(N \times 3n)$ of \mathbf{J}_μ is usually very large. So, we may use the approximation of \mathbf{J}_μ using initial $\nabla_{\mathbf{P}_0} \mathbf{I}$ and initial $f_{\mathbf{P}_0}$ at time t_0 . In this case, the approximation of the pseudo-inverse $\widehat{\mathbf{J}}_\mu^+$ is fixed and used instead of \mathbf{J}_μ^+ in (13).

III. VISUAL SERVOING WITH 4D ULTRASOUND

A 6-DOF robot equipped with a 4D US probe is used to compensate the motion of a target region in soft tissue structures. Therefore, only rigid motions such as 3D translations and 3D rotations are considered for the motion compensation. Nevertheless, a non-rigid motion model is necessary in order to track the deformable region. In this section, a rigid-motion extraction method and a 3D US position-based visual servoing approach are described.

A. Rigid Motion Extraction from Motion Parameters

The motion parameters of the non-rigid model above proposed are a set of control points. A rotation matrix \mathbf{R} and a translation vector \mathbf{t} describing the rigid motion are calculated using the following Least-Squares method with a set of initial control points \mathbf{C}_0 and a set of corresponding control points \mathbf{C} estimated by the non-rigid motion model. So, we will find \mathbf{R} and \mathbf{t} that minimize the following mean square error:

$$e^2(\mathbf{R}, \mathbf{t}, s) = \frac{1}{n} \sum_{i=1}^n \|\mathbf{c}_i - (s\mathbf{R}\mathbf{c}_i^0 + \mathbf{t})\|^2 \quad (14)$$

where s is a scale factor, n denotes the number of control points, and $\mathbf{c}_i^0 = (\tilde{x}_i^0, \tilde{y}_i^0, \tilde{z}_i^0)^T$ and $\mathbf{c}_i = (\tilde{x}_i, \tilde{y}_i, \tilde{z}_i)^T$ are an initial control point and its corresponding control point observed in the current 3D image. The solution of (14) is given by [10] [11] as explained below.

Let

$$\bar{\mathbf{c}}^0 = \frac{1}{n} \sum_{i=1}^n \mathbf{c}_i^0, \quad \bar{\mathbf{c}} = \frac{1}{n} \sum_{i=1}^n \mathbf{c}_i \quad (15a)$$

$$\sigma_{\mathbf{c}^0}^2 = \frac{1}{n} \sum_{i=1}^n \|\mathbf{c}_i^0 - \bar{\mathbf{c}}^0\|^2, \quad \sigma_{\mathbf{c}}^2 = \frac{1}{n} \sum_{i=1}^n \|\mathbf{c}_i - \bar{\mathbf{c}}\|^2 \quad (15b)$$

$$, \text{ and } \Sigma_{\mathbf{c}^0 \mathbf{c}} = \frac{1}{n} \sum_{i=1}^n (\mathbf{c}_i^0 - \bar{\mathbf{c}}^0)(\mathbf{c}_i - \bar{\mathbf{c}})^T. \quad (15c)$$

Now, we can calculate \mathbf{R} , \mathbf{t} , and s using a singular value decomposition of the covariance matrix $\Sigma_{\mathbf{c}^0 \mathbf{c}}$ expressed by $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.

$$\mathbf{R} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \quad \mathbf{t} = \bar{\mathbf{c}} - s\mathbf{R}\bar{\mathbf{c}}^0, \quad \text{and } s = \frac{1}{\sigma_{\mathbf{c}^0}^2} \text{tr}(\mathbf{\Sigma}\mathbf{S}) \quad (16)$$

$$\text{where } \mathbf{S} = \begin{cases} \mathbf{I} & , \text{ if } \det(\mathbf{U})\det(\mathbf{V}) = 1 \\ \text{diag}(1, 1, \dots, 1, -1) & , \text{ else } \end{cases}$$

B. 4D Ultrasound-Based Visual Servoing

We used a position-based visual servo control (PBVS) scheme described in [12] since the relative probe pose with respect to the soft tissue structures to track is calculated from the 3D rigid motion (\mathbf{R} and \mathbf{t}) of a target region extracted in a sequence of 3D US images. The sequential 3D US images are acquired from a 4D US probe that is mounted on the end-effector of a 6-DOF robot. We calculate the probe control velocity $\mathbf{v}_c = (\mathbf{v}_c, \mathbf{w}_c)$ of the 4D US probe according to the 3D rigid motion of the target region where \mathbf{v}_c and \mathbf{w}_c are the translational velocity vector and the angular velocity vector of the the probe frame.

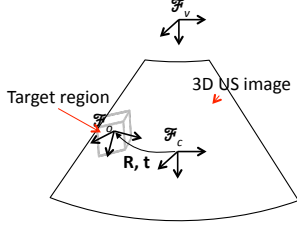


Fig. 2. Coordinate frames for position-based visual servo using 4D ultrasound

We set a probe frame \mathfrak{F}_c and an object frame \mathfrak{F}_o in the center of the initial target region and the current target region, respectively shown in Fig. 2. The objective of the PBVS is to move the probe to the center of the current target region in such a way to align the probe frame \mathfrak{F}_c on the object frame \mathfrak{F}_o .

The image feature \mathbf{s} is defined as $({}^{c^*}\mathbf{t}_c, \boldsymbol{\theta}\mathbf{u})$, in which ${}^{c^*}\mathbf{t}_c$ and $\boldsymbol{\theta}\mathbf{u}$ are a translation vector and the angle/axis parameterizations for the rotation matrix ${}^{c^*}\mathbf{R}_c$ which give the coordinate of the origin and the orientation of the probe frame \mathfrak{F}_c expressed in the desired probe frame \mathfrak{F}_{c^*} to achieve.

${}^{c^*}\mathbf{t}_c$ and ${}^{c^*}\mathbf{R}_c$ are given from the extracted rigid motion \mathbf{t} and \mathbf{R} of a target region as: ${}^{c^*}\mathbf{t}_c = -\mathbf{t}$, ${}^{c^*}\mathbf{R}_c = \mathbf{R}^{-1}$.

The strategy of the PBVS control scheme is to minimize the error $\mathbf{e}(t)$ between the current image feature ($\mathbf{s}(t) = ({}^{c^*}\mathbf{t}_c, \boldsymbol{\theta}\mathbf{u})$) and the desired image feature ($\mathbf{s}^* = \mathbf{0}$):

$$\mathbf{e}(t) = \mathbf{s}(t) - \mathbf{s}^* = \mathbf{s}(t)$$

The variation of \mathbf{s} with respect to the velocity of the probe is given by:

$$\dot{\mathbf{s}} = \mathbf{L}_s \mathbf{v}_c = \begin{bmatrix} {}^{c^*}\mathbf{R}_c & \mathbf{O} \\ \mathbf{O} & \mathbf{L}_{\theta\mathbf{u}} \end{bmatrix} \begin{bmatrix} \mathbf{v}_c \\ \mathbf{w}_c \end{bmatrix} \quad (17)$$

where \mathbf{L}_s is the 6×6 interaction matrix related to \mathbf{s} .

The time variation of the error \mathbf{e} is the same as the variation of \mathbf{s} : $\dot{\mathbf{e}} = \mathbf{L}_e \mathbf{v}_c$ where $\mathbf{L}_e = \mathbf{L}_s$. We apply the following control law to decrease the error \mathbf{e} exponentially, $\dot{\mathbf{e}} = -\lambda \mathbf{e}$.

$$\mathbf{v}_c = -\lambda \widehat{\mathbf{L}_e^{-1}} \mathbf{e} = -\lambda \begin{bmatrix} {}^{c^*}\mathbf{R}_c^\top & \mathbf{O} \\ \mathbf{O} & \mathbf{L}_{\theta\mathbf{u}}^{-1} \end{bmatrix} \begin{bmatrix} {}^{c^*}\mathbf{t}_c \\ \boldsymbol{\theta}\mathbf{u} \end{bmatrix} \quad (18)$$

Therefore,

$$\mathbf{v}_c = -\lambda {}^{c^*}\mathbf{R}_c^\top {}^{c^*}\mathbf{t}_c, \text{ and } \mathbf{w}_c = -\lambda \boldsymbol{\theta}\mathbf{u} \quad (19)$$

since $\mathbf{L}_{\theta\mathbf{u}}^{-1} \boldsymbol{\theta}\mathbf{u} = \boldsymbol{\theta}\mathbf{u}$.

Target motion component can be added to the above control law in order to enhance target tracking performance. In this case, the control law becomes:

$$\mathbf{v}_c = -\lambda \widehat{\mathbf{L}_e^{-1}} \mathbf{e} - \widehat{\mathbf{L}_e^{-1}} \frac{\partial \mathbf{e}}{\partial t}$$

where $\frac{\partial \mathbf{e}}{\partial t}$ is an approximation of the time variation of \mathbf{e} due to the target motion, that we estimate thanks to a Kalman filter

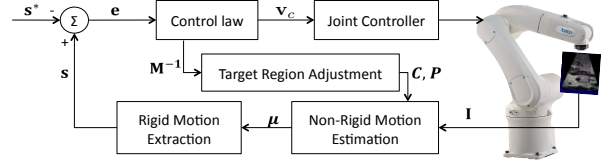


Fig. 3. 4D ultrasound-based visual servoing structure

which uses as input the relative pose of the object obtained during a sampling time Δt as follows:

$${}^{o(t)}\mathbf{M}_{o(t-\Delta t)} = {}^{o(t)}\mathbf{M}_{c(t)} \cdot {}^{c(t-\Delta t)}\mathbf{M}_{c(t)}^{-1} \cdot {}^{o(t-\Delta t)}\mathbf{M}_{c(t-\Delta t)}^{-1} \quad (20)$$

where ${}^a\mathbf{M}_b$ defines the homogeneous transformation matrix from frame b to frame a .

Note that the tracking target region \mathbf{I} should be adjusted every control loop as shown in the control scheme presented in Fig.3 because the robot compensates only the rigid motion part and the target region is not fully compensated. All the points \mathbf{P} in the target region and all the control points \mathbf{C} are adjusted according to the homogeneous matrix ${}^{v(t-\Delta t)}\mathbf{M}_{v(t)}$ describing the relative motion of the 3D US image frame \mathfrak{F}_v during time Δt :

$$\begin{bmatrix} \mathbf{c}_i(t) \\ 1 \end{bmatrix} = {}^{v(t-\Delta t)}\mathbf{M}_{v(t)}^{-1} \begin{bmatrix} \mathbf{c}_i(t-\Delta t) \\ 1 \end{bmatrix}, \forall i \in \{1, \dots, n\} \quad (21a)$$

$$\begin{bmatrix} \mathbf{p}_i(t) \\ 1 \end{bmatrix} = {}^{v(t-\Delta t)}\mathbf{M}_{v(t)}^{-1} \begin{bmatrix} \mathbf{p}_i(t-\Delta t) \\ 1 \end{bmatrix}, \forall i \in \{1, \dots, N\} \quad (21b)$$

IV. SIMULATION RESULTS

We simulated 4DUS in the presence of respiratory motion for the ground truth. The deformed images of a target region in 3D US images are generated using the non-rigid motion model above mentioned. We put a $19 \times 19 \times 19$ grid (called as a warping grid) in a warping region and $3 \times 3 \times 3$ control points (called as warping control points) in a target region of a 3D US image captured from an *in-vivo* human kidney as shown in Fig. 4. As each warping control point moves according to the respiratory motion model describe in the below section, all the vertices of the warping grid are relocated according to the motion model (6). Then, the 3D texture in each cell of the initial warping grid is mapped into the corresponding cell of the current warping grid.

The visual tracking algorithm was verified in a sequence of simulated 3D US images. We put control points for tracking at the same positions as the warping control points. The tracking results were assessed from the known positions of the warping control points. Furthermore, visual servoing was performed in the simulated environment to compensate the rigid motion.

Respiratory motion was modeled as (22) according to [5].

$$\mathbf{c}_i(t) = \mathbf{c}_i^0 + \mathbf{a} - \mathbf{b} \cos^{2n}(\pi t / \tau - \phi) + \boldsymbol{\eta}_i \quad (22)$$

where t , \mathbf{a} , \mathbf{b} , τ , ϕ , n , and $\boldsymbol{\eta}_i$ are the time in seconds, the position vector at inhale, the peak-to-peak extent vector of

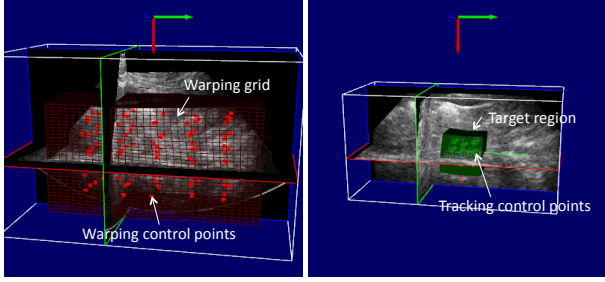


Fig. 4. Tracking of a target region (right) in a 3D US image deformed with a respiratory motion model (left)

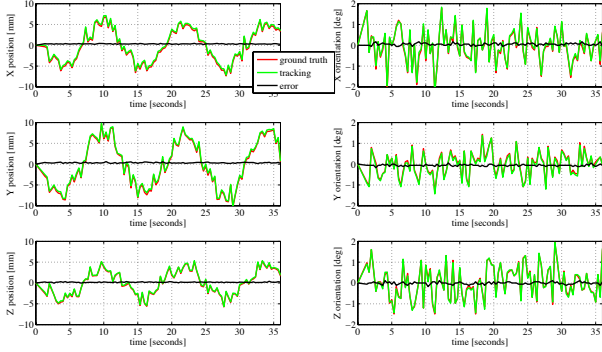


Fig. 5. Global rigid motion tracking results of a deformable region using simulated 4D US

motion, the cycle period, the starting phase of the breathing cycle, a flatness parameter of the model, and a noise-factor vector, respectively.

In this simulation, we used $n = 1$, $\tau = 12$ seconds, and $\phi = \pi/4$. The \mathbf{a} and \mathbf{b} are $(5, 7.5, 3.75)^T$ and $(10, 15, 7.5)^T$ in mm unit, respectively. The noise factor $\boldsymbol{\eta}_i$ was a vector of random numbers, $\eta_i^x \in [-1, 1]$, $\eta_i^y \in [-1.5, 1.5]$, $\eta_i^z \in [-0.75, 0.75]$.

We initialized a target region at the center of a 3D US image and the target region consists of $35 \times 35 \times 31$ points \mathbf{P}_0 and $3 \times 3 \times 3$ control points \mathbf{C}_0 (called as tracking control points) as shown in Fig. 4. While all the tracking control points follow the motions induced by all the warping control points, the image features s of a target region and a warping region are extracted. As a result, the tracking error was the mean absolute values of $(0.32, 0.33, 0.17, 0.05, 0.06, 0.04)$ in mm and degree units as we can see in Fig. 5. Furthermore, a servoing task to compensate the rigid motion was also performed successfully. The result are given in Fig. 6 and Fig. 7 with the maximum servo tracking error $\mathbf{e}_{\max} = (1.89, 3.28, 2.05, 0.74, 0.68, 1.03)$ in mm and degree units.

V. ROBOTIC EXPERIMENTS

A 4D US probe (model: 4DC7-3, Ultrasonix, Canada) was mounted on the end effector of a 6-DOF anthropomorphic robot to compensate the periodic motion induced by another robot as shown in Fig. 8. We used two types of TM phantoms as a target object, which are an US abdominal phantom (model: ABDGAN, Kyoto kagaku, Japan) and a deformable phantom made of gelatin by ourselves.

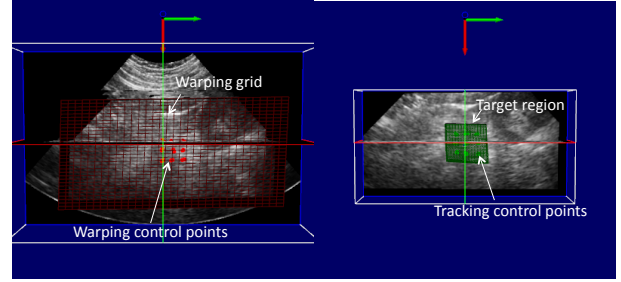


Fig. 6. Rigid motion compensation (right) in simulated environment (left)

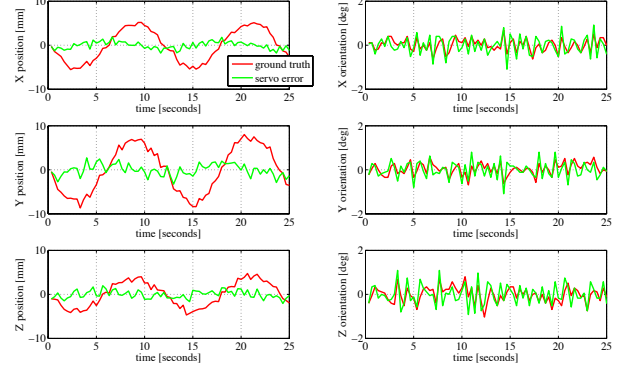


Fig. 7. Rigid motion compensation results of a deformable region using simulated 4D US

At first, we perform a rigid motion tracking task with an abdominal phantom. The secondary robot repeatedly rotated the abdominal phantom on a turning table in one direction and the opposite direction. In the meantime, the 6-DOF robot holding the 4D US probe is controlled by our method to automatically compensate the rigid motion of the target region within the phantom. The observed feature error and probe trajectory are shown in Fig. 9 and Fig. 10 (left). To maintain the firm contact between the probe and the phantom, we used a force control along the X axis of the 3D US image.

Secondly, a non-rigid motion tracking task was performed with a deformable phantom. The secondary robot repeatedly compresses/releases the deformable phantom laid in a drawer, and the 6-DOF robot conducted the tracking task as shown in Fig. 8 (right). Fig. 10 presents the evolution of the

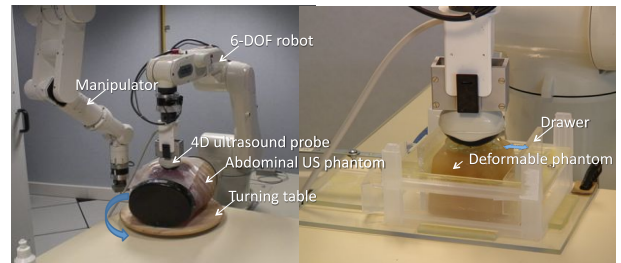


Fig. 8. A 6-DOF robot holds a 4D-US probe, and the other robot moves an abdominal phantom (left) and compresses/releases a deformable phantom (right)

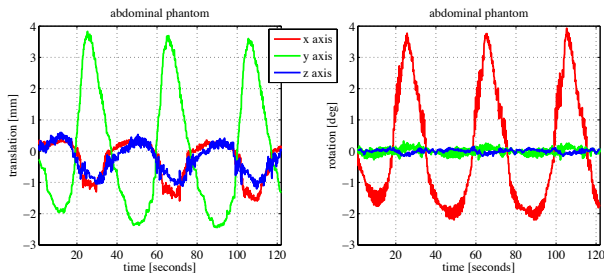


Fig. 9. Feature error that corresponds to the pose of a target region with respect to the current probe frame - case of the abdominal phantom

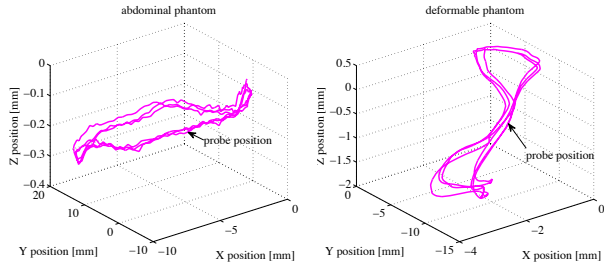


Fig. 10. Trajectories of the probe performed during the compensation visual servoing with an abdominal phantom (left) and a deformable phantom (right)

probe motion and Fig. 12 shows the observed feature error during the automatic compensation task performed with the deformable phantom. The visual tracking of the target region during the compression stage was successfully performed as shown in Fig. 11.

At the above experiments, sequential 3D US images were acquired at the rate of 4.6 volumes/second. The control loop time was 100 milliseconds. In order to perform all the processes explained in the section II and III such as 3D scan conversion, the non-rigid motion estimation and the rigid motion extraction in the control loop time (100 ms), we implemented them using nvidia CUDA.

VI. CONCLUSIONS

This paper has presented a method to compensate 3D non-rigid motion of soft tissue structures in the presence of respiration using 4D ultrasound. Motion parameters are estimated with respect to the changes of intensity values under the multiple tracking points within a target region. The rigid motion of the target region extracted from the motion parameters is compensated with a 6-DOF robot equipped with a 4D US probe. The non-rigid motion tracking

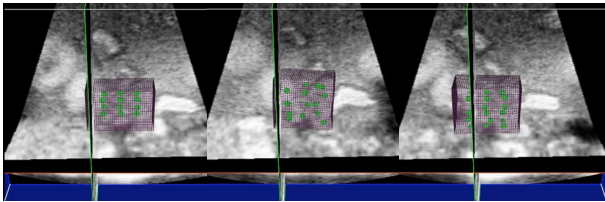


Fig. 11. Successive images obtained during the target region tracking process

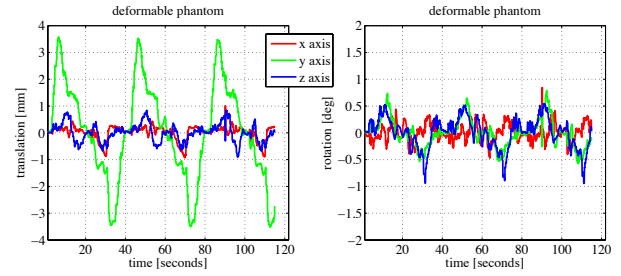


Fig. 12. Feature error that corresponds to the pose of a target region with respect to the current probe frame - case of a deformable phantom

method was validated in simulation by tracking the warping motion of an US volumetric image captured from *in-vivo* soft tissue. 4D US-based visual servoing tasks were performed successfully in the simulated deformations of a 3D US image. Furthermore, robotic experiments demonstrated non-rigid motion compensation with a deformable TM phantom in real time.

ACKNOWLEDGMENT

The authors acknowledge the support of the ANR project USComp of the French National Research Agency

REFERENCES

- [1] A. H. Gee, R. James Housden, P. Hassenpflug, G. M. Treece, and R. W. Prager, "Sensorless freehand 3D ultrasound in real tissue: speckle decorrelation without fully developed speckle." *Medical image analysis*, vol. 10, no. 2, pp. 137–49, Apr. 2006.
- [2] A. Krupa, G. Fichtinger, and G. D. Hager, "Real-time Motion Stabilization with B-mode Ultrasound Using Image Speckle Information and Visual Servoing," *The International Journal of Robotics Research*, vol. 28, no. 10, pp. 1334–1354, May 2009.
- [3] R. Mebarki, A. Krupa, and F. Chaumette, "2-D Ultrasound Probe Complete Guidance by Visual Servoing Using Image Moments," *IEEE Transactions on Robotics*, vol. 26, no. 2, pp. 296–306, Apr. 2010.
- [4] C. Nadeau and A. Krupa, "A multi-plane approach for ultrasound visual servoing : application to a registration task," *Ultrasound*, pp. 5706–5711, 2010.
- [5] E. J. Harris, N. R. Miller, J. C. Bamber, J. R. N. Symonds-Taylor, and P. M. Evans, "Speckle tracking in a phantom and feature-based tracking in liver in the presence of respiratory motion using 4D ultrasound." *Physics in medicine and biology*, vol. 55, no. 12, pp. 3363–80, June 2010.
- [6] F. L. Bookstein, "Principal Warps : Thin-Plate Splines and the Decomposition of Deformations," *Analysis*, vol. I, no. 6, 1989.
- [7] J. Lim, "A Direct Method for Modeling Non-Rigid Motion with Thin Plate Spline," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp. 1196–1202.
- [8] R. Richa and P. Poignet, "Three-dimensional Motion Tracking for Beating Heart Surgery Using a Thin-plate Spline Deformable Model," *The International Journal of Robotics Research*, vol. 29, no. 2-3, pp. 218–230, Dec. 2009.
- [9] G. Hager and P. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 10, pp. 1025–1039, 1998.
- [10] K. Arun, T. Huang, and S. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-9, no. 5, pp. 698–700, 1987.
- [11] S. Umeyama, "Least-Squares Estimation of Transformation Parameters Between Two Point Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376–380, 1991.
- [12] W. Wilson, C. W. Hulls, and G. Bell, "Relative end-effector control using cartesian position based visual servoing," *IEEE Transactions on Robotics & Automation*, vol. 12, pp. 684–696, 1996.