Intensity-based ultrasound visual servoing: modeling and validation with 2D and 3D probes

Caroline Nadeau, Alexandre Krupa

Intensity-based ultrasound visual servoing: modeling and validation with 2D and 3D probes

Caroline Nadeau, and Alexandre Krupa, *Member, IEEE,*

***Abstract*—In this work we present an ultrasound (US) visual servoing to control a robotic system equipped with a US probe. To avoid the difficult and time consuming image segmentation process, we develop a new approach taking as visual input directly the intensity of the image pixels. The analytic form of the interaction matrix that relates the variation of the intensity features to the motion of the probe is established and used to control the six degrees of freedom (dof) of the robotic system. Our approach is applied with a 2D and a 3D US probe and the results obtained with both sensors are compared in simulation. The 2D probe shows good performances for tracking tasks and the 3D one, that ensures a larger domain of convergence, is more particularly used for positioning tasks. The intensity-based approach is validated through experimental results performed with a realistic abdominal phantom and with animal soft tissues.**

***Index Terms*—Visual servoing, ultrasound, robotic system, intensity-based control.**

I. Introduction

A Large majority of image-guided robotic systems developed for surgical applications are based on a pre-operative planning of the gesture and the registration of this planning in the intra-operative environment. Such a strategy concerns for instance orthopedic or neurosurgical robots where a pre-operative planning defined by the surgeon is registered using fiducial markers fixed on the rigid surface of the bone or the skull. However with this pre-operative control, the gesture is reliable only as long as the patient and the anatomic target remain still. In particular in operations performed on soft tissues, this strategy is not robust to organs motions and deformations due to physiological motions. An alternative to this strategy is then a visual servoing approach where the robotic system is controlled in real-time using intra-operative information provided by a vision sensor. Among the different non-invasive medical imaging modalities, the US imaging is the only one that provides a visual feedback in real time using a non cumbersome transducer and a non ionizing energy, and that can therefore be used without any restriction for the patient as well as for the surgeon during a medical intervention. Because of such advantages, and despite the low quality of the image it produces, US is a promising imaging modality for image-guided robotic systems.

The previous works dealing with US visual servoing can be classified into two different system configurations. The former, namely eye-to-hand configuration, consists in controlling a surgical instrument using the visual feedback of a fixed US probe. The robotic manipulation offers a better accuracy than the human one and the proposed applications concern needle insertion procedures [3] or cardiac surgery [4]–[6]. The other configuration, namely eye-in-hand configuration, allows the direct control of the US sensor mounted on the robot end-effector for diagnostic purpose [7], [9] or surgical procedure [8].

In order to control one to six dof of the robotic manipulator, the efficiency of the visual servoing approaches are highly dependent on the choice of appropriate image features. Depending on the configuration, these features can be created by the intersection of the surgical tool with the US beam (eye-to-hand configuration) [3]–[6] or by anatomical landmarks (eye-in-hand configuration) [7]–[9].

In applications where a medical instrument is controlled in order to reach a target, both tool and target are segmented in the US image. In [3], the Hough transform is used to extract the axis of the instrument, rigidly aligned within the US probe plane and an active contour is manually initialized to track a target tumor. Two dof of a needle-insertion robot are then controlled by visual servoing to perform a percutaneous cholecystostomy while compensating involuntary patient motions. In the same way, in [4], a cross-shaped pattern is used as the target and a passive marker is fixed to the tool. Then a Radon transform is performed to extract these features in a 3D US image. However these transforms are specific for identifying long axes or detecting intersecting lines and can not be extended to detect all kinds of features. In [5], the four dof of a surgical forceps inserted in a beating heart through a trocar are controlled by visual servoing, using the two image points created by the intersection of the tool with the image plane and segmented using thresholding, morphological filtering and fast labeling process. In relation with this work, the authors of [6] developed a predictive control scheme to keep the forceps visible in the US image.

In robotic systems where the US probe itself is controlled, which are more particularly within the scope of this work, the image features can only be anatomic ones. In [7], five features extraction methods are compared to track an anatomical point corresponding to the center of an artery in order to servo the in-plane motions of the probe. These methods are based on image similarity measure such as cross correlation and sequential similarity detection or on contour segmentation by a Star [10] or Snake algorithm. For a lithotripsy procedure [8], which consists in the removal of kidney stones using high-intensity focused ultrasound (HIFU), two US probes and the

This work was presented in part at the IEEE Int. Conf. on Robotics and Automation, Shanghai, China, May 2011 [1] and at the IEEE Int. Conf. on Intelligent Robots and Systems, San Francisco, USA, September 2011 [2]. C. Nadeau is with Université de Rennes I, IRISA and INRIA Rennes-Bretagne Atlantique, Lagadic research group, 35042 Rennes, France. A. Krupa is with INRIA Rennes-Bretagne Atlantique and IRISA, Lagadic research group, 35042 Rennes, France. Alexandre.Krupa@inria.fr

HIFU transducer are mounted on the end effector of a XYZ stage robot to follow a target kidney stone while compensating physiological motions. The translational motions of the robotic effector are controlled with the 3D position of the kidney stone estimated from its segmentation in two orthogonal US images. Finally, approaches have been proposed to control the six dof of the probe with six geometric features built from 2D moments extracted from a single US image [9] or three orthogonal images [11]. However the moments computation requires a contour segmentation step whose efficiency depends on the organ shape and which is time consuming.

In this paper we propose a US visual servoing approach based directly on the image intensity for the control of both in-plane and out-of-plane motions of a US probe. The visual features involved in the control law are the intensities of a set of pixels contained in a fixed region of interest of the image. Contrarily to geometric features, the extraction of these intensities does not require any segmentation step and such a method can therefore deal with a large range of anatomic structures with no restriction due to the organ shape or the lack of contours.

Intensity-based features have been recently introduced with success in camera-based visual servoing. In this case, the image formation principle allows the modeling of the interaction between the time variation of these features and the six dof of the camera [12]. However, due to the specific geometry of the US sensor, which provides information only along its image plane, the control of the six dof of a 2D US probe requires additional out-of-plane information. In [13], Krupa *et al.* proposed an intensity-based approach to control a 2D US probe, using the speckle correlation observed in successive US images to control the out-of-plane motions of the probe. Nakadate *et al.* describe in [14] another intensity-based method to track the out-of-plane translation of the carotid artery. One dof of the robotic system is then controlled using an inter-frame block matching method to identify the artery motion. However both of these works are position-based and require a reconstruction of the pose of the robot effector from the image measurements.

On the contrary, we present here an image-based visual servoing where the robot control is directly performed in the image plane, which guarantees a good robustness to calibration errors. In particular we are able to compute the interaction matrix involved in the control law thanks to the 3D image gradient. Depending on the geometry of the considered US probe (2D or 3D) we develop different ways to compute this 3D gradient, either with derivative filters or with an on-line estimation algorithm. The 2D and 3D approaches are compared and analyzed in simulation environment for positioning and tracking applications and the first robotic results of this method on animal soft tissues are presented in this paper.

The structure of our paper is as follows. We firstly introduce the principle of the US visual servoing and insist on the characteristics of the US sensor. We then detail the intensity-based approach with the modeling of the interaction matrix. In Section IV, we focus on the computation of the 3D image gradient involved in the control law. Solutions to compute this gradient are proposed depending on the geometry of the considered US probe (2D or 3D) and an extensive simulation validation is realized to conclude on the advantages of each probe for tracking or positioning tasks. Finally, robotic experiments involving an hybrid force/vision control demonstrate the validity of the approach.

II. Ultrasound visual servoing

A. Image-based visual servoing

The principle of the image-based visual servoing consists in moving a robot so that a set of visual features $\mathbf{s}$ extracted from the image provided by a considered vision sensor reaches a set of desired features $\mathbf{s}^*$ observed at the desired pose $\mathbf{r}^*$ of the robot. The visual servoing control law is designed to minimize the visual error vector defined as $\mathbf{e}(t) = \mathbf{s}(t) - \mathbf{s}^*$. In order to try to ensure an exponential decoupled decrease of this error, the classical control law is given by [15]:

$$\mathbf{v}_p \;=\; -\lambda \, \widehat{\mathbf{L}_\mathbf{s}}^{+} \, (\mathbf{s}(t) - \mathbf{s}^*), \tag{1}$$

where λ is a positive gain, whose unit is s^{-1}, tuning the decrease of the visual error.

In an eye-in-hand configuration, $\mathbf{v}_p$ is the instantaneous velocity applied to the visual sensor and $\widehat{\mathbf{L}_\mathbf{s}}^{+}$ is the pseudo-inverse of an estimation of the interaction matrix $\mathbf{L}_\mathbf{s}$ that relates the variation of the visual features to the velocity $\mathbf{v}_p$ ($\dot{\mathbf{s}} = \mathbf{L}_\mathbf{s}\mathbf{v}_p$).

According to [15], the control scheme (1) is known to be locally asymptotically stable when a correct estimation $\widehat{\mathbf{L}_\mathbf{s}}$ of $\mathbf{L}_\mathbf{s}$ is used (i.e. as soon as $\widehat{\mathbf{L}_\mathbf{s}}^{+}\mathbf{L}_\mathbf{s} > 0$).

B. The US vision sensor

Traditionally, the visual servoing methods refer to vision data acquired with a camera mounted on a robotic system. In this case, the vision sensor provides a projection of the 3D world to a 2D image and a set of 2D image features can be used to control the six dof of the system. In the particular case of 2D US visual servoing, the image formation principle and the geometry of the vision sensor are far different from the ones of a camera.

1) US sensor model: With a 2D US probe, the created image corresponds to a cross-section of the visual target (see Fig. 1). Therefore, only the physical points lying in the US beam are represented in the US image.

Let the image plane, or probe plane, be defined by the beam of US waves emitted by the US sensor. The probe frame $\mathscr{F}_p$ is represented on Fig. 1, its origin is attached to the image center (u_0, v_0), $\mathbf{y_p}$ is aligned with the propagation direction of the US and $\mathbf{z_p}$ is orthogonal to the image plane. With this convention, the geometric model that relates the 3D coordinates of a physical point ${}^p\mathbf{X} = ({}^pX, {}^pY, {}^pZ)$ belonging to the probe plane (${}^pZ = 0$) with the pixel coordinates (u, v) of the corresponding image point is such as:

$$\left\{ \begin{array}{rcl} u &=& k_u \, {}^pX \;+\; u_0 \\ v &=& k_v \, {}^pY \;+\; v_0 \end{array} \right. , \tag{2}$$

where the scale factors (k_u, k_v) allow the conversion from metric coordinates to pixel coordinates. We can therefore define the respective scale factors $(s_x = 1/k_u, s_y = 1/k_v)$ that

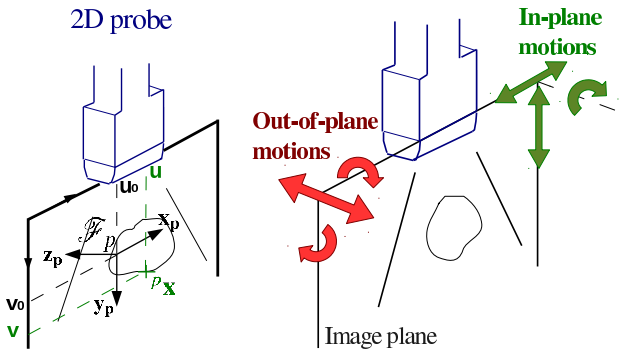


Fig. 1. Contrarily to a projective camera that gives a 2D projection of the 3D world, a 2D US probe provides visual information only within its image plane, making difficult the control of its out-of-plane motions.

convert the pixel coordinates to metric coordinates and correspond to the intrinsic parameters of the probe. With the extrinsic parameters of the probe that characterize its pose in a given reference frame, these intrinsic parameters are used to determine, for each image point, the 3D coordinates of the corresponding physical point. Both intrinsic and extrinsic parameters are estimated by a calibration procedure [16].

2) US wave physics: The US wave is a mechanical wave that generates pressure variations in the medium in which it travels. At the interface between two mediums of different impedances, the wave is partly transmitted and partly reflected towards the probe. The value of the reflected echo depends on the relative values of the acoustic impedances of both mediums. Moreover, when the incident ultrasonic wave is not orthogonal to the interface, both reflected and refracted waves are produced. The angle of the transmitted US beam depends on the angle of the incident beam and on the propagation velocity of the US in both mediums.

During their propagation in homogeneous materials, the US waves are also subject to two additional physics phenomena, the attenuation and the diffraction. The former depends on the distance to the US source and on an absorption linear coefficient of the medium. The latter occurs when the interface encounters by the US beam is small with respect to the US wavelength. The wave energy is then reflected in every direction, which creates a granular pattern in the US image called speckle.

3) B-mode image: The US signal can be easily represented by the amplitude of the echo reflected by interfaces it goes through. This representation mode, namely A-mode, allows the visualization of a single US line since the echo amplitude is given as a function of the distance to the source.

Currently, the more utilized representation of the US signal is the B-mode representation that associates to the US echo amplitude a grey level value (between 0 and 255). This representation allows the conservation of the spatial disposition of the US beams and the reconstructed B-mode image shows all US lines emitted by the probe. Compared to A-mode images, several processing are applied to the US signal, namely the attenuation compensation with an adaptive gain, the data interpolation to fill the full 2D image and a logarithmic compression of the signal.

Despite the physical characteristics of the US sensor, which is based on the propagation and reflection of acoustic waves, we treat it as a vision sensor since it provides gray-scale images in its common B-mode representation.

III. Intensity-based approach

The specificity of the US sensor in terms of geometry and image formation implies new challenges for the visual control, compared to camera-based approach. While light changes and depth estimation are no more an issue with the US modality, the major difficulties consist in the processing of the images and in the control of the out-of-plane motions of the sensor.

A. Intensity features

In this current work, we propose to address the first difficulty linked to the US image low quality and to the real-time constraint by avoiding any segmentation step and considering as visual features $\mathbf{s}$ the intensity values of the pixels of a region of interest (ROI) of the US image:

$$\mathbf{s} = (I_{1,1}, ..., I_{u,v}, ..., I_{M,N}), \tag{3}$$

where M and N are respectively the width and the height of the ROI and where $I_{u,v}$ represents the intensity of the pixel of 2D coordinates (u,v) in the US image.

B. Modeling of the interaction matrix

The second challenge is addressed by the computation of the interaction matrix $\mathbf{L_s}$ that links the variation of these intensity features to both in-plane and out-of-plane motion of the US probe in order to control the six dof of the probe. The objective is therefore to determine the analytic form of the time variation of one pixel intensity in the US image as a function of the probe velocity $\mathbf{v_c}$, that is $\mathbf{L}_{I_{u,v}}$ such as $\dot{I}_{u,v} = \mathbf{L}_{I_{u,v}}\mathbf{v}_p$.

Let the probe control frame $\mathscr{F}_p(\mathbf{x_p}, \mathbf{y_p}, \mathbf{z_p})$ be the frame attached to the center of the US image where $(\mathbf{x_p}, \mathbf{y_p})$ defines the image plane and $\mathbf{z_p}$ corresponds to the elevation axis, the 3D coordinates ${}^p\mathbf{x}_P = ({}^px_P, {}^py_P, {}^pz_P)$ of the point P, attached to the probe, that corresponds to the pixel p_x of coordinates (u,v) in the image are such as:

$$\begin{pmatrix} {}^px \\ {}^py \\ {}^pz \end{pmatrix} = \begin{pmatrix} s_x(u-u_0) \\ s_y(v-v_0) \\ 0 \end{pmatrix}, \tag{4}$$

where ${}^pz_P = 0$ since the point P belongs to the image plane.

The intensity $I_{u,v}(t)$ of the pixel p_x in the B-mode image acquired by the US probe at the time t corresponds to the amplitude $I_{US}(\mathbf{x}_O, t)$ of the US echo reflected by a physical point O of coordinates $\mathbf{x}_O$ belonging to the observed object:

$$I_{u,v}(t) = I_{US}(\mathbf{x}_O, t). \tag{5}$$

Under the hypothesis *(H1)* that *the probe is moving in a motionless environment*, at time $t+dt$ the point P, rigidly attached to the probe frame, coincides with a different physical point O' (see Fig. 2). The 3D coordinates of O' are $\mathbf{x}_O + \mathbf{dx}$, where $\mathbf{dx}$ is the displacement of P due to the probe motion

during the time dt. The intensity of the pixel p_x at time $t+dt$ is then equal to the US echo reflected by O':

$$I_{u,v}(t+dt) = I_{US}(\mathbf{x}_O + \mathbf{dx}, t+dt). \tag{6}$$

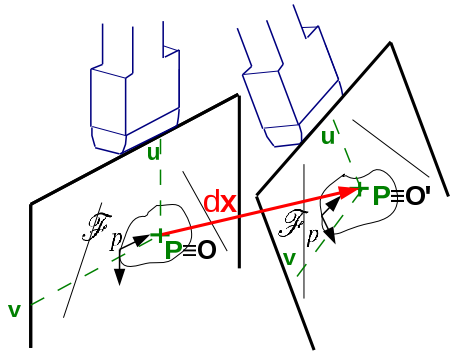


Fig. 2. Under the hypothesis of a moving probe in a motionless environment, the image point P, of constant coordinates (u,v) in the US image, coincides with a 3D point O at time t and with a different 3D point O' at time $t+dt$.

The computation of the interaction matrix is based on the hypothesis *(H2) of the constancy of the US wave reflection by a given physical structure*. Under this assumption, the US echo reflected by a physical point is independent of the time:

$$I_{US}(\mathbf{x}_O, t) = I_{US}(\mathbf{x}_O, t+dt) = I_{US}(\mathbf{x}_O).$$

This is a strong hypothesis since the intensity of the US echo reflected by an organ interface is dependent on the relative orientation of the US probe to this interface, but this conservation equation is *a posteriori* validated by the results obtained with an abdominal phantom and animal soft tissues (see Section VI). With *(H2)*, we can combine the equations (5) and (6) as follows:

$$I_{u,v}(t+dt) - I_{u,v}(t) = I_{US}(\mathbf{x}_O + \mathbf{dx}) - I_{US}(\mathbf{x}_O). \tag{7}$$

By approximating I_{US} by its first order Taylor polynomial around $\mathbf{x}_O$, it comes:

$$\begin{aligned} I_{US}(\mathbf{x}) \approx \quad & I_{US}(\mathbf{x}_O) + \frac{\partial I_{US}(\mathbf{x}_O)}{\partial x}(x-x_O) \\ & + \frac{\partial I_{US}(\mathbf{x}_O)}{\partial y}(y-y_O) + \frac{\partial I_{US}(\mathbf{x}_O)}{\partial z}(z-z_O). \end{aligned} \tag{8}$$

We set $\mathbf{x} = \mathbf{x}_O + \mathbf{dx}$ with $\mathbf{dx} = (dx, dy, dz)$ to obtain:

$$I_{u,v}(t+dt) - I_{u,v}(t) \approx \frac{\partial I_{US}}{\partial x}dx + \frac{\partial I_{US}}{\partial y}dy + \frac{\partial I_{US}}{\partial z}dz. \tag{9}$$

We divide (9) by dt to express the time variation of $I_{u,v}$ as a function of the motion ${}^p\dot{\mathbf{x}}_P$ of the image point P with respect to the environment and expressed in $\mathscr{F}_p$:

$$\dot{I}_{u,v} \approx \nabla I_{(u,v)}\ {}^p\dot{\mathbf{x}}_P, \tag{10}$$

with $\nabla I_{u,v} = [\nabla I_x\ \nabla I_y\ \nabla I_z]$ the 3D image gradient associated to the pixel (u,v). It is constituted of three components $\nabla I_x = \dfrac{\partial I}{\partial x}$, $\nabla I_y = \dfrac{\partial I}{\partial y}$ and $\nabla I_z = \dfrac{\partial I}{\partial z}$ that describe the intensity variation of the pixel (u,v) along the three axes of the image frame.

In the frame of the US probe $\mathscr{F}_p$, the velocity of the point P attached to the probe, with respect to the environment, is linked to the instantaneous velocity of the US probe $\mathbf{v_c} = ({}^p\boldsymbol{\nu}, {}^p\boldsymbol{\omega})$, according to the Varignon's formula of velocity composition in a solid [18]:

$${}^p\dot{\mathbf{x}}_P = ({}^p\boldsymbol{\nu} - [{}^p\mathbf{x}_P]_\times\ {}^p\boldsymbol{\omega}), \tag{11}$$

that can be rewritten as:

$${}^p\dot{\mathbf{x}}_P = \begin{bmatrix} \mathbf{I}_3 & -[{}^p\mathbf{x}_P]_\times \end{bmatrix} \mathbf{v}_p. \tag{12}$$

From (10), (12) and ${}^pz = 0$, the interaction matrix $L_{I_{u,v}}$ of size 1×6 associated to the visual feature $I_{u,v}$ is written as:

$$L_{I_{u,v}} = \begin{bmatrix} \nabla I_x & \nabla I_y & \nabla I_z & y\nabla I_z & -x\nabla I_z & x\nabla I_y - y\nabla I_x \end{bmatrix}, \tag{13}$$

and the complete interaction matrix $\mathbf{L_s}$ is built by stacking the $M\times N$ matrices $L_{I_{u,v}}$:

$$\mathbf{L_s} = \begin{pmatrix} L_{I_{1,1}} \\ \vdots \\ L_{I_{M,N}} \end{pmatrix}. \tag{14}$$

IV. Computation of the Image Gradient

To control the six dof of the US probe, the variation of the visual features is related to both in-plane and out-of-plane motions of the probe. In the interaction matrix, this variation is dependent on the 3D image gradient that has to be known. Solutions to compute this information are proposed here, depending on the geometry of the considered US probe.

A. With a classical 2D probe

1) 3D filter: In image processing, the image gradient is commonly estimated using directional image filters like the Sobel ones in the case of 2D gradient, that separately compute the horizontal and vertical derivatives of the image. These filters are based on two separable operators, a smoothing operator perpendicular to the derivative direction and a central difference in the derivative direction [18]. In the extended case of 3D image gradient, three filters can be designed on the same model to estimate the image gradient components $(\nabla I_x, \nabla I_y, \nabla I_z)$. Using two additional parallel images around the current one, a $3\times 3\times 3$ filter can be applied along each direction to compute these three components (see Fig. 3).

With a conventional 2D US probe, a small back and forth translational motion along the elevation direction is required to capture the additional images needed to compute the gradient components with the 3D filters. Because of this limitation, the gradient computation described above is dedicated more specifically to tracking tasks where the initial image is considered as the desired one and the visual task consists in the automatic stabilization of this image by compensating rigid motions of the target. During such a tracking task, the US probe remains close to its desired pose and the interaction matrix can be estimated once at the desired pose of the probe, with the back and forth motion, and no more updated during the servoing.

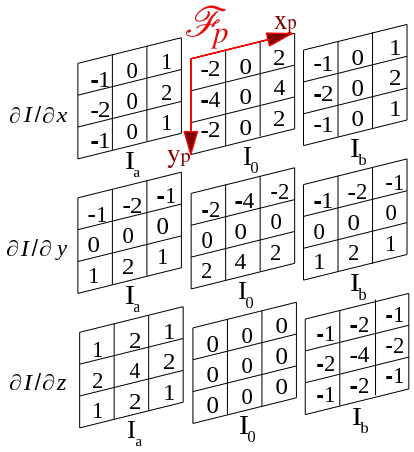


Fig. 3. The three filters are applied on each pixel of the current image I_0 to compute the gradient components $(\nabla I_x, \nabla I_y, \nabla I_z)$ using the additional parallel images I_a and I_b acquired on both sides of I_0.

2) On-line gradient estimation: We also propose an approach to estimate on-line the image gradient without additional motions of the probe to acquire parallel images. For this purpose we distinguish the in-plane components of the gradient $\nabla I_x, \nabla I_y$ from the out-of-plane component ∇I_z. The former are directly computed in the current US image with a 2D Sobel derivative filter while the latter is estimated from the current image and the previous ones acquired during the motion of the probe. This estimation is based on the approximation given by the Taylor expansion of the intensity of a 3D point of coordinates $\mathbf{x} = (x, y, z)$:

$$I_{US}(\mathbf{x}+\mathbf{dx}) \approx I_{US}(\mathbf{x}) + \nabla I_x\, dx + \nabla I_y\, dy + \nabla I_z\, dz. \tag{15}$$

Let $0(x, y, z)$ be the physical point coinciding with the image point P of pixel coordinates (u, v) at time t. If the probe moves, at time $t+dt$ the image point $P(u, v)$ coincides with a different physical point $0'(x', y', z')$. We have then the equality between the intensity of the US echo I_{US} reflected by O (resp. O') and the intensity of the pixel P of the US image acquired at time t (resp. $t+dt$):

$$\left\{ \begin{array}{rcl} I_{US}(x,y,z) & = & I_{u,v}(t) \\ I_{US}(x',y',z') & = & I_{u,v}(t+dt) \end{array} \right. . \tag{16}$$

We can express the coordinates of the physical points in the frame of the probe at time t referred to as $\mathscr{F}_p$:

$$\begin{pmatrix} {}^p x \\ {}^p y \\ {}^p z \end{pmatrix} = \begin{pmatrix} s_x\,(u-u_0) \\ s_y\,(v-v_0) \\ 0 \end{pmatrix} \tag{17}$$

and

$$\begin{pmatrix} {}^p x' \\ {}^p y' \\ {}^p z' \end{pmatrix} = {}^p\mathbf{R}_{p'} \begin{pmatrix} s_x\,(u-u_0) \\ s_y\,(v-v_0) \\ 0 \end{pmatrix} + {}^p\mathbf{t}_{p'}. \tag{18}$$

The rotation matrix ${}^p\mathbf{R}_{p'}$ and the translation vector ${}^p\mathbf{t}_{p'}$ describing the pose of the probe at time $t+dt$ expressed in the frame of the probe at time t are given by the robot odometry. Considering a small motion of the probe, O' is close to O and we can apply the equation (15):

$$\begin{array}{rcll} I_{US}({}^p x', {}^p y', {}^p z') & = & I_{US}({}^p x, {}^p y, 0) & +\nabla I_x\,({}^p x' - {}^p x) \\ & & +\nabla I_y\,({}^p y' - {}^p y) & +\nabla I_z\, {}^p z'. \end{array} \tag{19}$$

With the relationship (16), this equation can be rewritten as follows:

$$I_{u,v}(t+dt) = I_{u,v}(t) + \nabla I_x\,({}^p x' - {}^p x) + \nabla I_y\,({}^p y' - {}^p y) + \nabla I_z\, {}^p z'. \tag{20}$$

In this equation, the out-of-plane component of the gradient ∇I_z is related to a set of known values. The pixel intensities at time t and $t+dt$ and the in-plane gradient components, computed at time $t+dt$, are measured in the US image.

The equation (20) involves image measurements and is therefore sensitive to image noise. To ensure a better robustness of the gradient estimation, we implement a least squares method to compute ∇I_z from a set of several data.

The parametric model of the considered system (20) is rewritten in discrete time as:

$$Y_{[k+1]} = \Phi_{[k]}\ \theta_{[k]} + b_{[k+1]}, \tag{21}$$

whith:

$$\left\{ \begin{array}{rcl} Y_{[k+1]} & = & \left(I_{u,v[k+1]} - I_{u,v[k]} - \nabla I_{x[k]}\, dx_{[k]} - \nabla I_{y_{[k]}}\, dy_{[k]}\right) \\ \Phi_{[k]} & = & dz_{[k]} \\ \theta_{[k]} & = & \nabla I_{z_{[k]}} \end{array} \right. . \tag{22}$$

θ is the parameter of the system to estimate, Y and Φ are measured at each iteration and since we have no a priori on the model noise we choose b as a white noise term to represent the image noises and robot measure errors.

With the parametric representation, the output of the system can be *a priori* predicted $\widehat{Y}_{[k+1]}$ from the parameter estimate $\widehat{\theta}_{[k]}$:

$$\widehat{Y}_{[k+1]} = \Phi_{[k]}\ \widehat{\theta}_{[k]}.$$

The *a priori* error between this prediction and the real value of the system output, $E_{[k+1]}$ is then defined as:

$$\begin{array}{rcl} E_{[k+1]} & = & Y_{[k+1]} - \widehat{Y}_{[k+1]} \\ & = & Y_{[k+1]} - \Phi_{[k]}\ \widehat{\theta}_{[k]} \end{array} \tag{23}$$

The aim of the least squares approach is to estimate the parameter θ that minimizes a criterion defined as the quadratic sum of the prediction errors $E_{[k+1]}$ obtained at each iteration:

$$J(\hat{\theta}_{[k]}) = \sum_{j=0}^{k} \beta^{k-j} \left(Y_{[j]} - \Phi_{[j]}\, \widehat{\theta}_{[k]}\right)^2. \tag{24}$$

The scalar β $(0 < \beta < 1)$ is a weighting factor, also called forgetting factor, used to lower the importance of past data.

Several approaches have been proposed to solve this minimization, among which we can distinguish non recursive strategies from recursive ones.

Indeed, the limitations of the recursive algorithms are underlined in [19]. The stability of these algorithms is not guaranteed without the introduction of a dead zone or a stabilization term when considering noises and disturbances. In the same time, the improvement brought by these modifications in terms of robustness is done at the expense of the algorithm efficiency.

Finally, when dealing with noised signals, direct identification methods are shown to offer better robustness and results [19].

We propose therefore to consider a Sliding Least Squares algorithm, where the estimate of the parameter $\hat{\theta}_{[k]}$ is computed using a set of measures Y and Φ acquired on a window of size N_{LS}. This estimate is defined in [19] for the multi-dimensional case and can be written as follows for a mono-dimensional problem:

$$\hat{\theta}_{[k]} = \begin{cases} Q_{[k]}/R_{[k]} & \text{if} \quad R_{[k]} > \varepsilon_0 \\ \hat{\theta}_{[k-1]} & \text{if} \quad R_{[k]} \leq \varepsilon_0 \end{cases}, \tag{25}$$

with:

$$\begin{aligned} R_{[k]} &= \sum_{j=k-N_{LS}+1}^{k} \left(\beta^{(k-j)} \Phi_{[j]}^T \Phi_{[j]} \right) \\ Q_{[k]} &= \sum_{j=k-N_{LS}+1}^{k} \left(\beta^{(k-j)} \Phi_{[j]} Y_{[j]} \right) \end{aligned} \tag{26}$$

A threshold value ε_0 is defined to fix a boundary on the smallest value of $R_{[k]}$ that can be taken into account. ε_0 guarantees that enough out-of-plane motion has been performed between two successive iterations of the algorithm to update the out-of-plane gradient information. ε_0 is such as $\varepsilon_0 = \delta_0^2$ where δ_0 is the estimated minimal distance from which the intensity variation of one pixel is relevant with respect to the image noise specific to the US sensor. Typically we chose δ_0 about the size of one pixel.

B. The benefits of the 3D probes

For the intensity-based visual servoing, the advantage of the 3D probes is to directly provide out-of-plane information that can be used in the visual features vector or to compute the interaction matrix. Depending on the technology of the 3D probe, two different approaches are considered.

1) 3D approach: A matrix array 3D probe provides a 3D US volume in real-time. In this case, the ROI considered in the visual servoing strategy is a set of voxels, whose intensities are used as visual information:

$$\mathbf{s} = (I_{1,1,1}, ..., I_{u,v,w}, ..., I_{M,N,L}),$$

where M, N and L are respectively the width, height and depth of the volume. With the modeling of the interaction matrix described in section III.B, we can write the new interaction matrix associated to the intensity of one voxel $V(u,v,w)$ as:

$$\mathbf{L}_{I_{u,v,w}} = \nabla I_{(u,v,w)} \begin{bmatrix} \mathbf{I}_3 & -[{}^p\mathbf{x}_V]_\times \end{bmatrix}, \tag{27}$$

with $\nabla I_{(u,v,w)}$ the 3D image gradient of the voxel V and ${}^p\mathbf{x}_V = (x,y,z)$ the coordinates of V in the probe frame. Given (u_0, v_0, w_0) the voxel coordinates of the probe frame origin and s_z the voxel depth size, we have:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} s_x(u-u_0) \\ s_y(v-v_0) \\ s_z(w-w_0) \end{pmatrix}. \tag{28}$$

In particular z is no more equal to zero and the 1×6 interaction matrix associated to the visual feature $I_{u,v,w}$ is then:

$$L_{I_{u,v,w}} = [\nabla I_x \;\; \nabla I_y \;\; \nabla I_z \;\; y\nabla I_z - z\nabla I_y \;\; -x\nabla I_z + z\nabla I_x \;\; x\nabla I_y - y\nabla I_x], \tag{29}$$

Using the complete volume provided by the 3D probe, the current interaction matrix can be computed on-line by filtering this volume with the 3D derivative filters previously described (see Fig. 3).

2) 2D approach: The current limitation of the matrix array probes is their small field of view (FOV) and the low quality of the obtained volume. On the contrary, motorized 3D probes offer a good quality of images but a trade-off is required between the size of the out-of-plane FOV and the probe frame rate. For instance, with a motor step of 1.4 deg, a volume of 28 deg FOV is reconstructed from a set of 20 images at a frame rate of 3 vol/sec. To increase this frame rate and perform real-time control of the probe, we propose to consider a 2D approach. The visual information is extracted from one image plane and the 3D information is only used for the image gradient computation. In this case, we do not need to capture a complete 3D US volume but only 2 or 4 additional images acquired with the motorized sweeping of the 2D transducer (see Fig. 4) at a frame rate of 12 vol/sec.

Since a motorized 3D probe does not acquire parallel images due to the rotation angle α (see Fig. 4), new 3D derivative filters are proposed in this section to take into account the elevation distance between the image pixels.

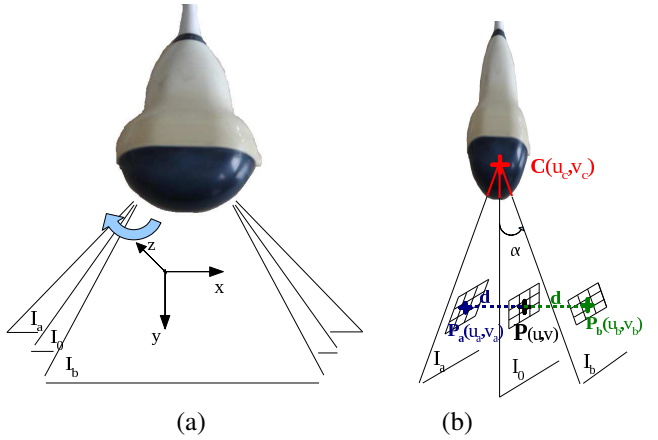


Fig. 4. (a) The additional US images provided by the 3D probe are no more parallel to the current one but are obtained after a rotation of an angle α around the image x-axis. (b) The derivative filters are weighted to take into account the variable distance d.

Given $C(u_c, v_c)$ the point corresponding to the US beams intersection. The 3D filter applied to the pixel $P(u,v)$ of the current image I_0 corresponds to a set of three 2D filters respectively applied to $P_a(u_a, v_a)$, $P(u,v)$ and $P_b(u_b, v_b)$ where P_a and P_b are the image points of I_a and I_b that orthogonally project onto P in I_0. Their coordinates are computed from the coordinates of P and the angle α:

$$\begin{cases} u_a = u_b = u \\ v_a = v_b = \dfrac{(v+v_c)}{cos(\alpha)} - v_c \end{cases} \tag{30}$$

With the respective orientations of the acquired US images, the elevation distance between P and P_a or P and P_b is dependent on the pixel ordinate v in the image. Weights inversely proportional to this elevation distance are defined

as follows:

$$w(v) = \frac{1}{d(v)} = \frac{1}{(v+v_c)\,tan(\alpha)}. \quad (31)$$

These weights are multiplied to the coefficients applied to the images I_a and I_b with the non-weighted derivative filters (see Fig. 3).

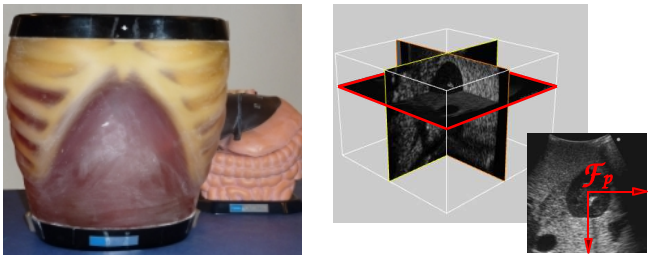
V. Simulation validation

To validate the US intensity-based approach, we use a software simulator that we have developed to reconstruct and display a dense volume from a set of parallel images. Positioning and tracking tasks are then performed in this simulation environment to assess the advantages and the limits of our approach, respectively using 2D and 3D probes.

A. Simulation environment

The simulator is built from the Visualization ToolKit (VTK) library [20] and the Visual Servoing Platform (ViSP) [21], both being open source C++ libraries. The US simulator provides an external view of the loaded US volume by means of two central orthogonal slices as shown in Fig. 5. Moreover, in addition to this display functionality, the simulator allows the control of a virtual US probe and generates the internal view of this probe by a cubic interpolation process.

For the simulation validation, we use the US volume of a kidney acquired on a realistic abdominal phantom (see Fig. 5). This volume is created from a set of 335 parallel images of size 250×250 and pixel size of $0.6 \times 0.6 mm^2$, which were automatically acquired using elevation intervals of 0.3 mm.



(a) (b)

Fig. 5. (a) The US abdominal phantom AB-41900-030 Kyoto Kagaku - ABDFAN. (b) The volume loaded in the simulator is represented by two orthogonal slices and the virtual probe plane, defined with the frame $\mathcal{F}_p$, is displayed in red.

B. Positioning tasks

We first simulate a positioning task, using the simulation environment to obtain a ground truth of the evolution of the pose error of the US probe. We position the virtual probe in the simulator and we consider the corresponding image as the desired one I^*. Then the probe is moved away to a new pose where the observed organ section is considered as the initial image and the visual servoing is launched. In the pose expression, the $\theta\mathbf{u}$ representation is considered to describe the orientation, where $\mathbf{u} = (u_x\, u_y\, u_z)^\top$ is a unit vector representing the rotation axis and θ is the rotation angle.

The results of several positioning tasks are gathered in Fig. 6. The same control law (1) is applied to the probe with a gain $\lambda = 1$, empirically adjusted to ensure the best behavior of the control law, and different methods for the computation of the 3D gradient are compared.

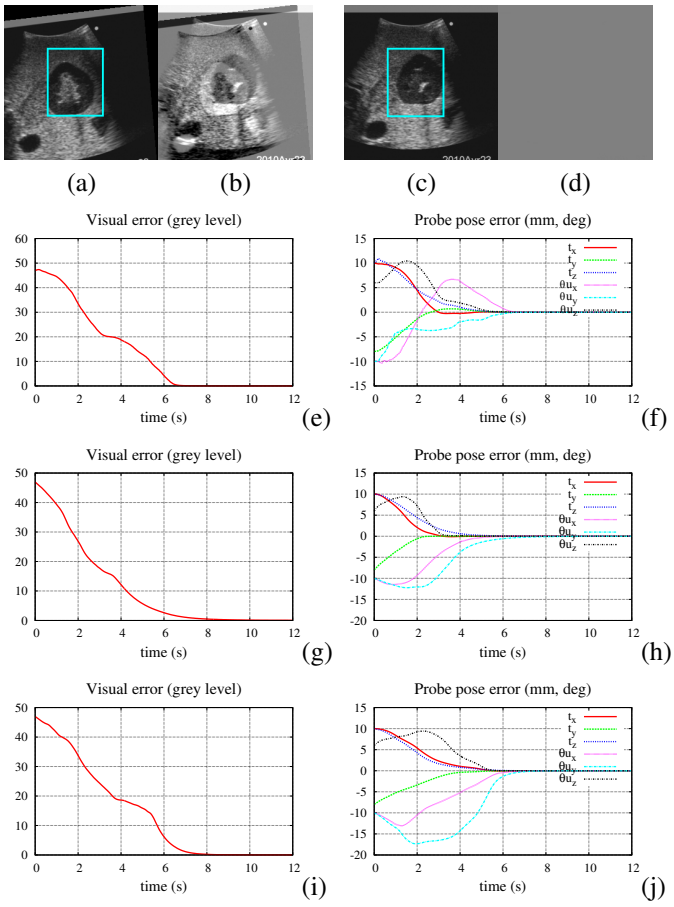


Fig. 6. The visual convergence from the initial US view (a) to the final one (c) is shown by the corresponding difference image with the desired view (respectively (b) and (d)). Under the same simulation conditions, we compare the different solutions proposed to compute the image gradient with the non-recursive estimation algorithm (e,f) and with the weighted (g,h) and non weighted (i,j) derivative filters. These latter curves correspond to the results obtained with a 3D probe (IV.D.2).

The view of the virtual probe is shown at its initial (a) and final (c) positions. The cyan rectangle defines the ROI of size 100×150 pixels. To display the visual error between the current I and desired I^* images during the positioning task, a difference image I_{diff} is computed as:

$$I_{diff} = \frac{(I - I^*) + 255}{2}.$$

The difference images corresponding to the initial view of the probe and to the final one are given in Fig. 6(b) and (d). The uniform gray color of this difference image after the convergence of the algorithm demonstrates the success of the positioning task since the final image perfectly coincides with the desired one. Moreover, we define a visual error function $\mathscr{C}$ to visualize the evolution of the error between the current

and desired features vectors during the visual servoing:

$$\mathscr{C} = \sqrt{\frac{(\mathbf{s}-\mathbf{s}^*)^\top (\mathbf{s}-\mathbf{s}^*)}{N_{pix}}}.$$

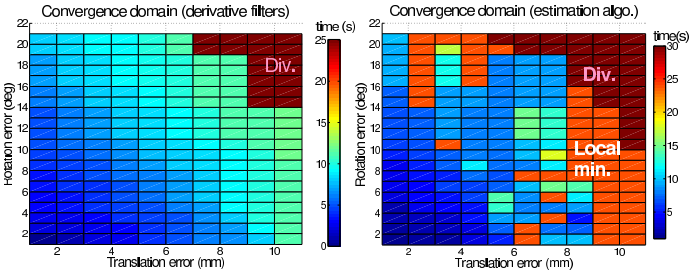
$\mathscr{C}$ is expressed in grey levels and normalized by the number of pixels N_{pix} in the considered ROI.

1) On-line gradient estimation: In this positioning task, the in-plane gradient components ∇I_x and ∇I_y are computed in the current image with 2D derivative filters and the out-of-plane component is estimated with the on-line algorithm. During the five first iterations of the algorithm, an open loop translation is applied to the probe to initialize the estimation algorithms. Then the value of the parameter ∇I_z is updated each time enough out-of-plane motion is applied to the probe. The results of the positioning task, presented in Fig. 6, show the visual convergence (e) and the pose convergence (f) of the control law. From an initial error $\Delta\mathbf{r}(mm/deg) = (10,-8,10,-10,-10,6)$, the final pose error of the probe is less than 0.1 mm in translation and 0.1 deg in rotation.

2) Derivative gradient filters: The curves (g) and (h) correspond to the use of a virtual motorized 3D probe. In this case, the visual information is extracted from one image plane and the 3D geometry allows the acquisition of additional images to compute on-line the 3D image gradient with derivative filters. The virtual 3D probe provides at each iteration one current image along its plane and four additional images tilted with an angle $\pm\alpha = \pm 1.4$ deg and $\pm 2\alpha = \pm 2.8$ deg. The 3D image gradient is directly computed from this set of images with three weighted $5\times5\times5$ filters, which avoids the initial out-of-plane motion of the probe required with the 2D approach. The performance of the weighted derivative filters we have designed for a set of five non parallel images is shown by comparison with the results obtained with the non-weighted filters (curves (i) and (j)). Even if the desired pose is reached in both cases, with the same gain of the control law, the convergence is faster and more direct with the weighted filters.

3) Robustness to calibration errors and images noises: After the validation of the control law under ideal conditions, we introduce systemic and random errors likely to occur in the real robotic system. We take into account a calibration error between the robotic end-effector and the image frame of $5deg$ on each rotation and $5mm$ on each translation. We also add normal Gaussian noises on the pixel intensities and on the pose measures with a variance of respectively 3 grey levels for the intensity and $3mm$ and $1deg$ for the position and orientation measures. Under these simulation settings and considering the same desired pose of the probe, several positioning tasks are launched from different initial poses of the US probe in order to assess the convergence domain of the positioning task. The two methods of gradient computation (with derivative filters and with the estimation algorithm) are compared in Fig. 7 where the time of convergence is expressed as a function of the initial error of the probe.

As expected, with the gradient computation by image filters, the visual servoing approach is very robust to small calibration errors. Logically, the on-line estimation of the gradient is more sensitive to calibration errors and local minimums are regularly

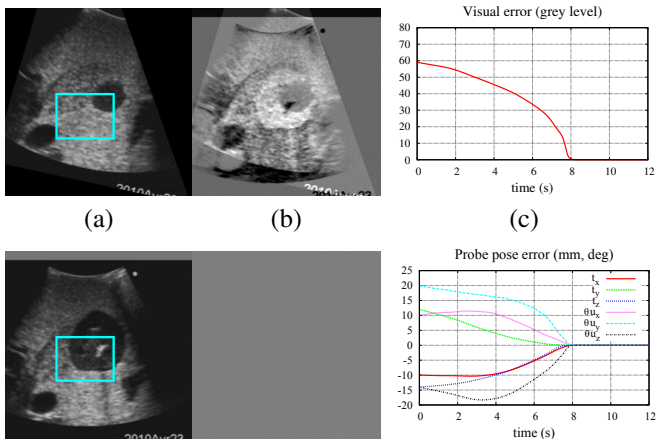


(a) (b)

Fig. 7. The convergence domain of the intensity-based approach depends on the method of 3D gradient computation. With the derivative filters (a) the convergence time increases proportionally with the distance to the desired probe pose and the divergence (in brown) is observed for an initial translation superior to $10mm$ and an initial rotation superior to $14deg$. With the estimation algorithm, local minimums (in orange) are often reached as soon as the initial pose is not very close to the desired one.

reached in this case when the initial pose error of the probe is superior to $5mm$ for each translation and $8deg$ for each rotation.

4) 3D visual information for a larger convergence domain: In the previous simulations only 2D information, which is more generally available with US probes, was considered in the visual vector. However the development of matrix-array 3D probes makes now possible the use of 3D visual information in the control law according to the strategy presented in section IV-B1. A new positioning task is presented in Fig. 8 where 20 parallel images are acquired in the simulation environment on both sides of the displayed view.



(a) (b) (c)

(d) (e) (f)

Fig. 8. Positioning task with a 3D probe, using visual features extracted from a 3D volume. The central view of the volume is displayed at the initial (a) and final (d) probe poses with the corresponding difference images (b,e). The visual (c) and pose (f) convergence is displayed during the task.

A volume of $80\times60\times41$ voxels is then considered in the control law and filtered with the 3D derivative filters (see Fig. 3) to compute the interaction matrix. To reach the same desired pose of the probe as previously, a different initial pose is chosen here, farther than in the 2D simulations and from which the one-plane algorithm falls in a local minimum:

$$\Delta r_{init}(mm,deg) = (-10,12,-14,10,20,-14).$$

With the 3D approach, the local minimum is avoided and the

positioning task is well performed. We can therefore note that the 3D information provided by the virtual 3D probe ensures a larger domain of convergence of the visual servoing.

C. Tracking tasks

The second robotic task we target is the active stabilization of a US image. In this case, the initial image is also the desired one, and the probe is moved to compensate for external motions applied to the observed object. In the following simulations, a sinusoidal motion of period $5s$ and amplitude $Ay = 7mm$ along the vertical translation and $Az = 15mm$ along the out-of-plane translation is applied to the US volume of the kidney that has been loaded in the simulator. In parallel, the velocity $\mathbf{v}_c$ computed from the intensity-based control law (1) with $\lambda = 1.5$ is applied to the US probe and the results obtained respectively with the current and desired interaction matrices are presented in Fig. 9.

1) with on-line estimation of the image gradient: After the open-loop motion of the probe required to initialize the estimation algorithm, the current interaction matrix is computed with the image gradient components estimated by the non recursive estimation algorithm. In order to compute the variation of the parameter ∇I_z due to the probe motion and not to the disturbance motion, we set the least squares window size to a small value $N_{LS} = 3$ and the forgetting factor to $\beta = 0.9 < 1$. This temporal window size corresponds to a duration of $120ms$ that is extremely short with respect to the $5s$ of the disturbance period. By setting these parameters, the US volume can therefore be considered static during the $120ms$ of the data acquisition. The stabilization is accurately performed since the maximum visual error is 4 grey levels (c) and the maximum pose error is $0.2mm$ (d).

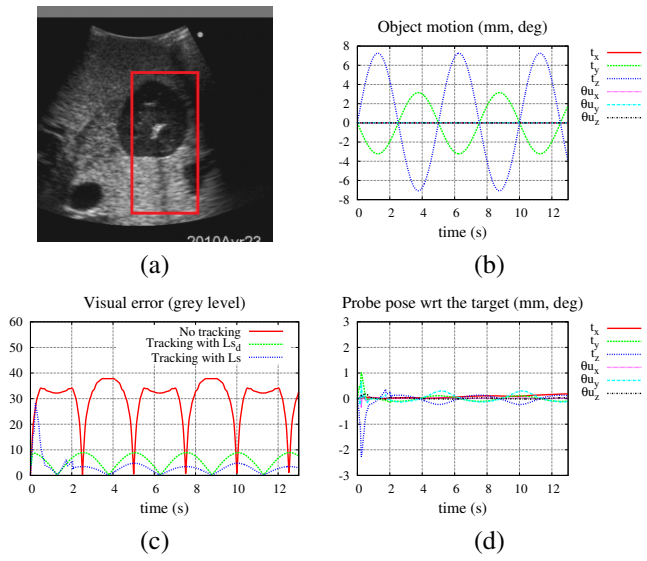


Fig. 9. Tracking task with a 2D probe. Stabilization of the desired US image (a) while the kidney volume undergoes a translational periodic motion (b). (c) The visual error is highly reduced with both methods, which corresponds to a low pose error (here displayed for the use of $\mathbf{L_s}$) (d).

We can note from Fig. 10 that the gradient is not updated continuously during the tracking. When the probe is perfectly synchronized with physiological motions (around $t = 4s$ for instance), the current and desired visual features are very close and the velocity computed by the visual servoing control law, that depends to the visual error, is low. Therefore the displacement is not sufficient to update the image gradient because of the threshold of eq. (25). On the contrary, when some delay appears (around 5s and 10s), the probe velocity increases with the image error and the gradient image can be estimated.

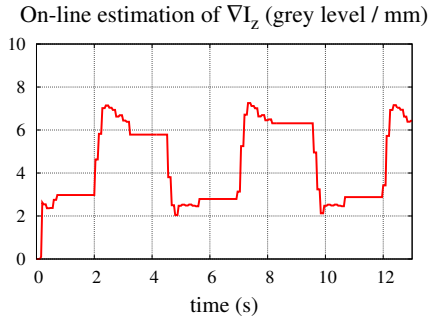


Fig. 10. Evolution of the out-of-plane value of the image gradient for one pixel of the ROI during the tracking task.

2) with the desired interaction matrix: The current pose of the probe being always close to its desired pose in this tracking task, the interaction matrix is well approximated by the desired interaction matrix [15]. This matrix, referred to as $\mathbf{L_{s^*}}$, is computed once and for all at the initial pose without being updated during the servoing task. For this computation, five parallel images are acquired around the desired pose of the US probe and the 3D derivative filters presented in Fig. 3 are applied to this set of images. The results obtained with $\mathbf{L_{s^*}}$ are compared to the ones obtained with $\mathbf{L_s}$ in terms of visual error (c). The pose error is very similar to the previous simulation, which validates the use of $\mathbf{L_{s^*}}$ in this particular case.

D. Conclusion

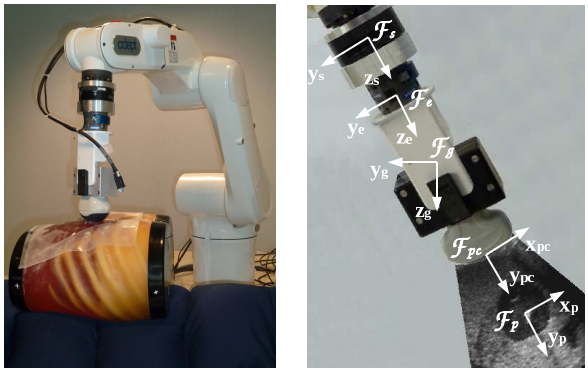
Based on simulation validations, the 2D intensity-based approach shows good results for local positioning tasks but is generally sensitive to local minimums when considering further initial poses of the probe. In this case, the 3D probes offer a better robustness to calibration errors in the computation of the interaction matrix and give more accurate results for positioning tasks from a remote initial pose, thanks to an increased convergence domain.

On the contrary, the 2D probes are more fitted for tracking tasks than the 3D ones thanks to their higher frame rate. If the tracking is accurate enough and the probe remains always close to the desired US view, the current interaction matrix is well approximated by the desired one $\mathbf{L_{s^*}}$, that can be computed at the initial pose of the probe before the appearance of the disturbance. In the perspective of the breathing motion compensation, $\mathbf{L_{s^*}}$ could be estimated while the patient holds his breath since this step requires no more than few seconds. However if the dynamics of the system is not sufficient to track accurately the disturbance motion, then the on-line estimation

of the current interaction matrix $\mathbf{L_s}$ is necessary to compensate for the tracking delays (see Section VI-C).

VI. Robotic experiments

Experiments have been performed on a realistic US abdominal phantom, using an anthropomorphic robotic arm equipped with a US transducer and a force sensor (see Fig. 11(a)).



(a) (b)

Fig. 11. (a) The ADEPT Viper robotic system with a motorized 3D US probe. (b) The robot end effector (frame $\mathcal{F}_e$) is equipped with a force sensor (frame $\mathcal{F}_s$) and a 2D US probe (frame $\mathcal{F}_p$).

A. Hybrid vision/force control

In the following experiments we combine the visual control with a force control since the US probe is in contact with the surface of the phantom. Two sensors are then involved in an hybrid vision/force control based on an external control loop approach [22]. The force control is dedicated to the control of the translational motion along the y-axis of the probe frame while the five remaining dof are controlled by visual servoing.

1) Force control: We implement a force control law to guarantee a constant resulting force of $1N$ applied on the contact point pc of the probe with the object surface along the y-axis of the probe frame. ${}^{pc}H_{pc}$ corresponds to the contact force tensor expressed in the frame $\mathcal{F}_{pc}$, which is centered on the contact point and aligned with the probe frame $\mathcal{F}_p$ (see Fig. 11(b)). This tensor is expressed as:

$$ {}^{pc}H_{pc} = {}^{pc}\mathbf{F}_s \left({}^{s}H_s - {}^{s}\mathbf{F}_g \, {}^{g}H_g \right) \tag{32} $$

where ${}^{a}\mathbf{F}_b$ is a transformation matrix used to express in the frame $\mathcal{F}_a$ a force tensor known in the frame $\mathcal{F}_b$:

$$ {}^{a}\mathbf{F}_b = \begin{bmatrix} {}^{a}\mathbf{R}_b & \mathbf{0}_{3\times3} \\ [{}^{a}\mathbf{t}_b]_\times \, {}^{a}\mathbf{R}_b & {}^{a}\mathbf{R}_b \end{bmatrix} \tag{33} $$

${}^{a}\mathbf{t}_b$ and ${}^{a}\mathbf{R}_b$ are the translation vector and the rotation matrix of the frame $\mathcal{F}_b$ with respect to the frame $\mathcal{F}_a$ and $[{}^{a}\mathbf{t}_b]_\times$ is the skew symmetric matrix related to ${}^{a}\mathbf{t}_b$.

${}^{s}H_s$ is the total force tensor measured by the force sensor and ${}^{s}\mathbf{F}_g \, {}^{g}H_g$ is the gravity force applied to the force sensor due to the mass m_p of the US probe, both are expressed in the force sensor frame. The ${}^{g}H_g$ tensor is defined as ${}^{g}H_g = [0\ 0\ 9.81m_p\ 0\ 0\ 0]^T$ in the frame $\mathcal{F}_g$ centered on the mass center of the probe as indicated in Fig. 11(b).

We express then the resulting force tensor in the probe frame $\mathcal{F}_p$ and we compute the instantaneous velocity of the probe $\mathbf{v}_f$ generated by the following proportional force control law:

$$ \mathbf{v}_f = -\frac{K \, {}^{p}\mathbf{F}_{pc} \left({}^{pc}H_{pc} - {}^{pc}H_{pc}{}^{*} \right)}{k}, \tag{34} $$

where ${}^{pc}H_{pc}{}^{*} = [0\ 1N\ 0\ 0\ 0\ 0]^T$ is the desired contact force, k is an estimate of the contact stiffness and K is the control gain.

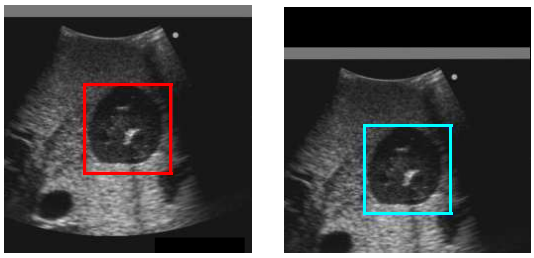
2) Vision/force fusion: We use a 6×6 selection matrix $\mathbf{M_s} = diag(0,1,0,0,0,0)$ to apply the force control only along the y-axis of the probe. The complementary matrix $(\mathbf{I}_6 - \mathbf{M_s})$ is then introduced to apply the vision control on the five remaining dof of the probe. To combine the force and the vision control, we send the following angular velocity $\dot{\mathbf{q}}$ to the end effector of the robotic arm:

$$ \dot{\mathbf{q}} = {}_{e}\mathbf{J}_e{}^{-1} \mathbf{v}_e = {}_{e}\mathbf{J}_e{}^{-1} \, {}^{e}\mathbf{W}_p (\mathbf{M_s} \, \mathbf{v}_f + (\mathbf{I}_6 - \mathbf{M_s}) \, \mathbf{v}_p), \tag{35} $$

where ${}_{e}\mathbf{J}_e{}^{-1}$ is the inverse of the robot Jacobian and where ${}^{e}\mathbf{W}_p$ is the transformation matrix that transforms a velocity skew from the probe frame to the effector frame:

$$ {}^{e}\mathbf{W}_p = \begin{bmatrix} {}^{e}\mathbf{R}_p & [{}^{e}\mathbf{t}_p]_\times \, {}^{e}\mathbf{R}_p \\ \mathbf{0}_{3\times3} & {}^{e}\mathbf{R}_p \end{bmatrix} \tag{36} $$

As we choose for safety reasons to give priority to the force control over the vision control, the latter can fail to converge to the desired image since the translational velocity component along the y-axis computed by the visual servoing control law is not applied to the probe. To deal with this issue, we apply this velocity to the ROI itself to readapt its position in the US image (see Fig. 12).



(a) (b)

Fig. 12. Principle of the combination of the vision and force controls. (a) A target US slice with in red the desired ROI. (b) To oppose the vertical displacement of the probe due to the force control, the vision control is applied to the ROI, which is translated inside the US image.

B. Positioning task with a 3D probe

Up to now, the matrix-array 3D probes have a small field of view and provide a volume of low resolution, particularly difficult to process. Therefore the positioning task is here performed with a 3D motorized probe (4DC7-3/40, Ultrasonix) with a frequency bandwidth of 3 to 7 MHz dedicated for abdominal imaging, using the 2D approach detailed in Section IV.B.2. With a depth of $12cm$ and a motor step of 1.4 deg, this probe allows the acquisition of a small volume constituted of three images at a frame rate of $12vol/sec$. The control loop time is defined by this volume frame rate and at each iteration

the 3D image gradient is computed with the 3D derivative filters. The gain of the control law is set to $\lambda = 0.4$ and the results obtained with the abdominal phantom are given in Fig. 13.

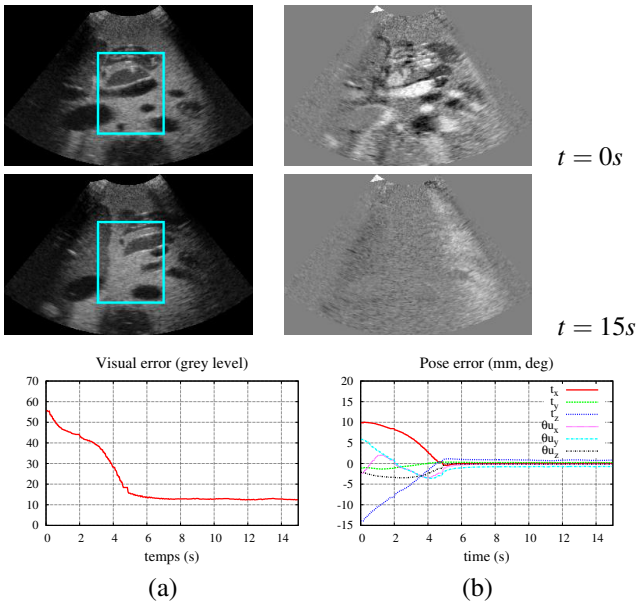


Fig. 13. Results of a positioning task with a 3D probe. The central slice acquired by the probe is displayed at the beginning and the end of the task. The visual error (a) does not reach zero at convergence because of the noise of the sensor but the task is validated in terms of pose error (b).

The internal view of the probe and the difference image are shown at the initial ($t = 0s$) and final ($t = 15s$) poses of the 3D probe. Only the central image acquired by the probe is displayed, the additional ones being only used to compute the 3D image gradient. The visual convergence of the task can be seen on the uniform difference image corresponding to the final pose of the probe and on the curve (a) that corresponds to the evolution of the visual error during the positioning task. The pose convergence is also observed (see Fig. 13(b)) and the initial pose error: $\Delta r_{init}(mm, deg) = (9.9, -1.1, -13.9, -2.4, 5.8, -2.1)$ is minimized to: $\Delta r_{fin}(mm, deg) = (0.1, 0.2, 0.7, -0.1, -0.7, -0.1)$.

C. Tracking task of a fast and large motion

The robotic arm is now equipped with a 2D convex US probe of 2-5 MHz frequency bandwidth (C60, Sonosite) to perform a tracking task. We position the 2D US probe on the abdominal phantom and we define the ROI in the US image (see Fig. 14(a)). The force/vision control is launched after a small automatic back and forth out-of-plane translation used to initialize the estimation of the 3D image gradient. Then we manually apply various large and fast translational and rotational motions to the phantom. The dynamics of this disturbance exceeds the one of the control law in order to create important delays in the tracking and assess the ability of the control to overcome these delays. The tracking results are shown in Fig. 14.

In the current experiment where important probe pose errors appear due to the tracking delay, we update the interaction

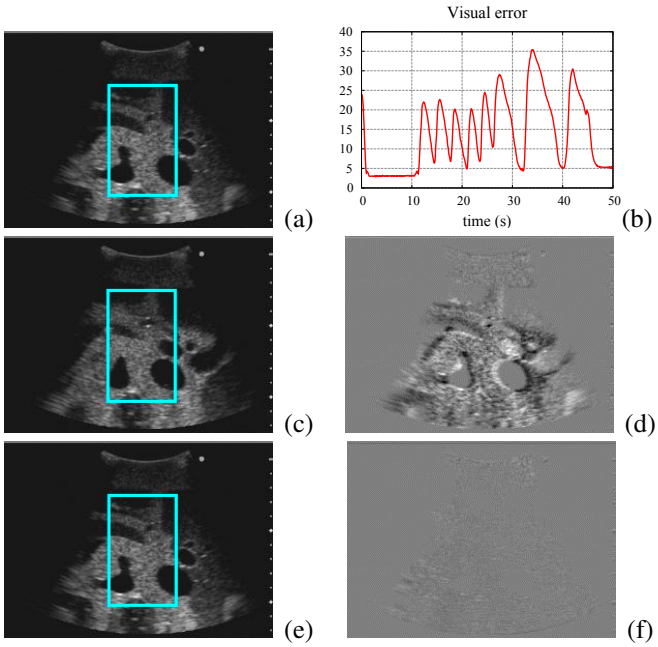


Fig. 14. Tracking of an abdominal cross-section (a) with a 2D probe. The tracking delay of the 2D probe is observed through the evolution of the visual error (b) and through the US image (c) and difference image (d) corresponding to the maximum visual error (at $t = 33s$). Nevertheless, at the end of the disturbance, the probe reaches its desired pose as is shown by the final US image (e) and difference image (f).

matrix during the tracking task thanks to the on-line estimation of the image gradient.

The interaction matrix, initialized at the desired pose, is then computed with the image gradient extracted directly from the current image (for the 2D components ∇I_x and ∇I_y) and estimated from the image measures as described in Section IV-A2 (for its out-of-plane component ∇I_z). Despite the important disturbances applied to the phantom, the US probe follows the phantom motion and converges to the desired image when the motion stops.

D. Tracking task with a 2D probe: first ex-vivo results

For this validation, a chicken stuffed with pig liver and kidneys (Fig. 15(a)) and immersed in a water tank to avoid air gaps inside its body is carried by a 6 dof robot (Robot2). A periodic motion is applied to this phantom with Robot2 while the tracking task is performed with a second robotic arm (Robot1) equipped with a 2D US probe. The probe frame rate is 25 images/s and the control loop time is $40ms$. Two optical markers are fixed on the probe and on the phantom and provide the relative pose of both elements thanks to an EasyTrack system. This relative pose is only used as a ground truth to validate the tracking task.

Robot1 is manually positioned above the phantom and the force control is applied with a force of 3N to put the probe in contact with the chicken surface. A desired ROI is then defined in the US image and a small back and forth motion is automatically realized to compute $\mathbf{L}_{\mathbf{s}^*}$ with a set of parallel images acquired around the desired image. A 3D periodic motion along all translations and one rotation (around the probe axis) is then applied to the phantom. This disturbance

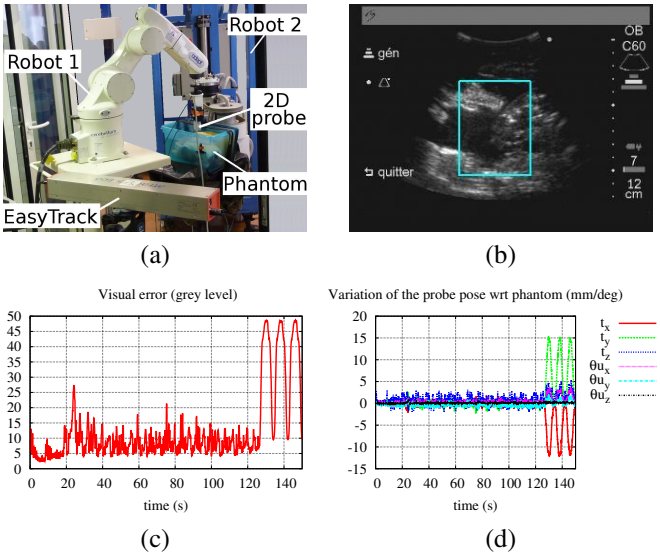


Fig. 15. Ex-vivo results: experimental setup (a) and desired US image to compensate (b). The tracking, launched at $t = 0s$ and stopped at $t = 125s$, ensures the compensation of the disturbance applied for $t > 20s$ as can be seen on the visual error (c) and the pose error (d).

has a period of $8s$ and generates amplitudes of motion of $12mm$ and $15mm$ along horizontal axes and $10mm$ along the vertical axis with a rotation of the phantom of $4deg$. To increase the tracking dynamics, we implement a Kalman filter to predict the target motion. The Kalman filter is based on a constant velocity model and takes as input the measures of the image features variation and the probe instantaneous velocity to provide an estimate of the target velocity. This estimated velocity $\hat{\mathbf{v}}_\mathbf{o}$ is finally reinjected into the control law (1) as a prediction term:

$$\mathbf{v}_p = -\lambda \, \mathbf{L}_{\mathbf{s}^*}^{+} \left(\mathbf{s}(t) - \mathbf{s}^*\right) + \hat{\mathbf{v}}_\mathbf{o}.$$

The results of one tracking task with the 2D probe are displayed in Fig. 15 for a gain $\lambda = 0.4$. The disturbance motion is applied at $t = 20s$, then at $t = 125s$ the compensation is stopped. The curve (c) shows the minimization of the visual error throughout the tracking task and the curve (d), obtained thanks to the EasyTrack system, validates the robotic task in terms of pose since the relative pose of the probe with respect to the phantom is maintained constant during the tracking ($t < 125s$).

In this experiment, the introduction of the Kalman filter in the control law improves the accuracy of the tracking task for sinusoidal motions with smooth changes of directions. For more complex motions, other predictive controller should be considered, such as for instance a repetitive predictive controller R-GPC [23] if the period of the disturbance is known.

VII. Conclusion and Perspectives

In this paper a new approach is proposed to control all the motions of a US probe by visual servoing. The originality of the approach is due to the direct use of the US B-mode image as visual information. In particular no image processing or segmentation step is required to build the visual features vector that corresponds here to the intensity values of the pixels contained in a ROI of the current US image. The interaction matrix associated to these visual features is modeled and the proposed intensity-based approach is validated in simulation and with several robotic experiments where the US probe interacts with an abdominal phantom and animal soft tissues. 2D and 3D probes geometries are successively considered to perform positioning and tracking tasks. The 2D approach shows good results for local positioning tasks and especially tracking tasks thanks to the high frame rate of the 2D probe while the 3D probes offer a better robustness to calibration errors and give more accurate results for positioning tasks from a further initial pose.

So far no in-vivo validation of the intensity-based approach has been performed and the current work is a first step that introduces and validates this new US visual servoing in experimental conditions, with slow and rigid motions to compensate. A further step will be to target a specific medical procedure to increase the robustness of the approach and deal with in-vivo constraints. In this context, several medical applications have been identified that could benefit from the proposed approach.

With a tele-echography robot, the automatic positioning of the probe to a pre-recorded US image and its tracking during the tele-operation can ease the diagnosis of the user. The aim is to virtually stabilize an organ of interest with a robotic task based on the fusion of the vision and tele-operation controls. More generally, the active stabilization of an anatomic cross-section could improve the diagnosis, based on US wave intensity measures, blood flow measures (Doppler) or enhanced imaging, that requires the observation of a stable anatomic target during several minutes and can be compromised by any motion of the patient. For the same reason, the measure of arteries dilation induced by blood flow, known as "Flow Mediated Dilation" [24] can benefit from the stabilization of the US image. Finally, surgical applications, such as the destruction of a kidney stone by high intensity focused US or the irradiation of tumors are also targeted. The visual servoing approach can be used to follow the target and then maximize its irradiation while avoiding the healthy tissues.

Caroline Nadeau Caroline Nadeau received the M.S. degree in electrical engineering and automation from the National Polytechnic Institute of Toulouse, France, in 2008 and the Ph.D. degree in signal processing from the University of Rennes, France, in 2011. Her Ph.D. research work on medical robotics and visual servoing from ultrasound images was carried out with the Lagadic group at IRISA - Inria Rennes-Bretagne Atlantique, France. Since 2012, she has been a research scientist with CEA List, France, where she is currently a member of the Laboratoire Images, Tomographie et Traitement. Her current research interests include image-based robotics control, image processing and computed tomography.

Alexandre Krupa Alexandre Krupa received the M.S. and Ph.D. degrees in control systems and signal processing from the National Polytechnic Institute of Lorraine, Nancy, France, in 1999 and 2003, respectively. He received the "Habilitation Diriger des Recherches" in Signal processing from the University of Rennes, France, in 2012. His Ph.D. research work was carried out with the eAVR team (Control Vision and Robotics) with the Laboratoire des Sciences de l'Image de l'Informatique et de la Télédétection, Strasbourg, France. From 2002 to 2004, he was an Assistant Associate Professor for undergraduate student lectures in electronics, control, and computer programming with Strasbourg University, Strasbourg, France. Since 2004, he has been a Research Scientist with Inria Rennes, France, where he is currently a member of the Lagadic group. In 2006, he was a Postdoctoral Associate with the Computer Integrated Surgical Systems and Technology Engineering Research Center, Johns Hopkins University, Baltimore, MD. His current research interests include medical robotics, computer-assisted systems in the medical and surgical fields, and, more specifically, the control of medical robots by visual servoing using ultrasound images.

ACKNOWLEDGMENT

The authors acknowledge the support of the ANR project US-Comp of the French National Research Agency and thank Philippe Poignet and Nabil Zemiti for making available the robotic platform of LIRMM to conduct the experiments on animal soft tissues.

REFERENCES

[1] C. Nadeau, A. Krupa, Intensity-based direct visual servoing of an ultrasound probe. *IEEE Int. Conf. on Robotics and Automation, ICRA'11*, Shanghai, China, May 2011.

[2] C. Nadeau, A. Krupa, Improving ultrasound intensity-based visual servoing: tracking and positioning tasks with 2D and bi-plane probes. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'11*, San Francisco, USA, September 2011.

[3] J. Hong, T. Dohi, M. Hashizume, K. Konishi, N. Hata, A motion adaptable needle placement instrument based on tumor specific ultrasonic image segmentation. *5th Int. Conf. on Medical Image Computing and Computer Assisted Intervention, MICCAI'02*, pp. 122-129, Tokyo, Japan, September 2002.

[4] P.M. Novotny, J.A. Stoll, P.E. Dupont and R.D. Howe, Real-time visual servoing of a robot using three-dimensional ultrasound. *IEEE Int. Conf. on Robotics and Automation, ICRA'07*, pp. 2655-2660, Roma, Italy, April 2007.

[5] M.A. Vitrani, H. Mitterhofer, N. Bonnet, G. Morel, Robust ultrasound-based visual servoing for beating heart intracardiac surgery. *IEEE Int. Conf. on Robotics and Automation, ICRA'07*, pp. 3021-3027, Roma, Italy, April 2007.

[6] M. Sauvee, P. Poignet, E. Dombre, US image based visual servoing of a surgical instrument through non-linear model predictive control. *Int. Journal of Robotics Research*, vol. 27(1), January 2008.

[7] P. Abolmaesumi, S. Salcudean, W. Zhu, M. Sirouspour, and S. DiMaio, Image-guided control of a robot for medical ultrasound. *IEEE Trans. on Robotics*, vol. 18(1), February 2002.

[8] D. Lee, N. Koizumi, K. Ota, S. Yoshizawa, A. Ito, Y. Kaneko, Y. Matsumoto, and M. Mitsuishi, Ultrasound-based visual servoing system for lithotripsy. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'07*, pp. 877-882, 2007.

[9] R.Mebarki, A. Krupa and F. Chaumette, 2D ultrasound probe complete guidance by visual servoing using image moments. *IEEE Trans. on Robotics*, vol. 26(2): 296-306, 2010.

[10] N. Friedland and D. Adam, Automatic ventricular cavity boundary detection from sequential ultrasound images using simulated annealing. *IEEE Trans. Med. Imag.*, vol. 8(4): 344-353, 1989.

[11] C. Nadeau, A. Krupa, A multi-plane approach for ultrasound visual servoing: application to a registration task. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'10*, Taipei, Taiwan, October 2010.

[12] C. Collewet, E. Marchand, Photometric visual servoing. *IEEE Trans. on Rob.*, 27(4):828-834, August 2011.

[13] A. Krupa, G. Fichtinger, G. Hager, Real time motion stabilization with B-mode ultrasound using image speckle information and visual servoing. *Int. Journal of Robotics Research*, vol. 28(10):1334-1354, 2009.

[14] R. Nakadate, J. Solis, A. Takanishi, E. Minagawa, M. Sugawara, K. Niki, Out-of-plane visual servoing method for tracking the carotid artery with a robot-assisted ultrasound diagnostic system.*IEEE Int. Conf. on Robotics and Automation, ICRA'11*, Shanghai, China, May 2011.

[15] F. Chaumette, and S. Hutchinson, Visual servo control, Part I: Basic approaches. *IEEE Robotics and Automation Magazine*, vol. 13(4): 82-90, 2006.

[16] L. Mercier, T. Lango, F. Lindseth, and L.D. Collins, A review of calibration techniques for freehand 3-D ultrasound systems. *Ultrasound in Medicine and Biology*, vol. 31(2): 143-165, 2005.

[17] V. Acary, B. Brogliato, Numerical Methods for Nonsmooth Dynamical Systems: Applications in Mechanics and Electronics (Springer Verlag), pp. 114, 2008.

[18] M. Hadwiger, J. M. Kniss, C. Rezk-salama, D. Weiskopf, and K. Engel, Real-time Volume Graphics. pp.112-114, 2006.

[19] M. de Mathelin, and R. Lozano, Robust adaptive identification of slowly time-varying parameters with bounded disturbances. *Automatica*, vol. 35(7): 1291-1305, 1999.

[20] W. Schroeder, K. Martin, and B. Lorensen, The Visualization Toolkit: An Object Oriented Approach to 3D Graphics. *3rd Edition, ISBN-1930934076, Kitware, Inc. Publisher*, 2003.

[21] E. Marchand, F. Spindler, F. Chaumette. ViSP for visual servoing: a generic software platform with a wide class of robot control skills. *IEEE Robotics and Automation Magazine*, vol. 12(4): 40-52, December 2005.

[22] J. De Schutter, T. De Laet, J. Rutgeerts, W. Decr, R. Smits, E. Aertbelin, K. Claes, and H. Bruyninckx, Constraint-based task specification and estimation for sensor-based robot systems in the presence of geometric uncertainty. *Int. Journal of Robotics Research*, vol. 26(5): 433-455, 2007.

[23] J. Gangloff, R. Ginhoux, M. De Mathelin, L. Soler, and J. Marescaux, Model predictive control for compensation of cyclic organ motions in teleoperated laparoscopic surgery. *IEEE Trans. on Control Systems Technology*, vol. 14(2), 2006.

[24] M. C. Corretti, T. J. Anderson, E. J. Benjamin, D. Celermajer, F. Charbonneau, M. A. Creager, J. Deanfield, H. Drexler, M. Gehard-Herman, D. Herrington, P. Vallance, J. Vita, R. Vogel, Guidelines for the Ultrasound Assessment of Endothelial-Dependent Flow-Mediated Vasodilation of the Brachial Artery. *Journal of American College of Cardiology*, vol. 39(2): 257-265, 2002.