



# Improving ultrasound intensity-based visual servoing: tracking and positioning tasks with 2D and bi-plane probes

C. Nadeau, A. Krupa

## ► To cite this version:

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, 2011, San Francisco, USA, United States. pp.2837-2842. hal-00639692

**HAL Id: hal-00639692**

**<https://inria.hal.science/hal-00639692>**

Submitted on 9 Nov 2011

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

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

# Improving ultrasound intensity-based visual servoing: tracking and positioning tasks with 2D and bi-plane probes

Caroline Nadeau, Alexandre Krupa

**Abstract**—Real time and non invasive, the ultrasound imaging modality can easily be used in minimally invasive surgery or needle insertion procedures to visualize an organ or a tumor to reach. However the manual stabilization of the ultrasound image while the organ moves with patient breathing or heart beating can be very tricky. In this paper, we present an intensity-based approach to control both in-plane and out-of-plane motions of an ultrasound probe held by a robotic arm in order to reach and follow one organ cross section. Two methods are proposed to improve the accuracy of this intensity-based approach, by estimating on-line the 3D image gradient required in the control law and by considering a bi-plane sensor. Robotic experiments are performed with two different ultrasound sensors on a realistic abdominal phantom and validate this visual servoing approach.

**Index Terms**—Visual servoing, intensity-based control, ultrasound, bi-plane sensor

## I. INTRODUCTION

The visual servoing techniques consist in integrating the visual feedback provided by a vision sensor in order to control the movements of a dynamic system with a closed-loop approach [1]. Historically, a large majority of the visual servoing approaches proposed in the literature consider a monocular camera to control a robotic system. More particularly in the medical field, the first image-guided robotic systems have been introduced in the context of laparoscopic surgery, using as vision sensor an endoscopic camera [2]. Artificial markers fixed on a surgical tool were used as visual features to control the endoscope in order to keep the instrument centered in the image.

In the particular case of ultrasound (US) visual servoing, the image formation principle and the geometry of the vision sensor which gives information only in its image plane are far different from the ones of a camera. Consequently, the challenges in the visual control differ. While light changes and depth estimation are no more an issue with the US modality, the major challenges consist in the processing of the images and in the control of the out-of-plane motions of the sensor. First works in US visual servoing focus on the control of the in-plane motions of the probe. In [3], the three in-plane degrees of freedom (dof) of a robotic system are controlled to maintain the section of an artery centered in the US image during a manual out-of-plane translation of the US probe. In [4], two dof of a needle-insertion robot are controlled by visual servoing to perform a percutaneous cholecystostomy

while compensating involuntary patient motions. The target and the needle are automatically segmented in the US images and their respective poses are used to control the robot.

Some authors have proposed solutions to control the out-of-plane motions of the probe by visual servoing. In [5], an actuated surgical tool is guided to a desired target under 3D US imaging. Both instrument and target are segmented and tracked in the 3D image and the pose error is then used to control 4 dof of the instrument. However 3D US devices currently provide small and low-quality volumes at a low acquisition rate which limits their use in real-time robotic applications. In another work [6], the three translations of a XYZ stage robot equipped with two US probes and a HIFU transducer are controlled to follow a kidney stone while compensating for physiological motions during a lithotripsy procedure. With a 2D US probe, Vitrani *et al.* [7] proposed to use the two image points corresponding to the intersection of a surgical forceps with the probe plane as visual features to control 4 dof of the tool inside a beating heart in an eye-to-hand configuration. In relation with this work, the authors of [8] developed a predictive control scheme to keep the forceps visible in the US image. More recently, different approaches have been proposed to control the 6 dof of the probe with six geometric features built from 2D moments extracted from a single US image [9] or three orthogonal images [10]. However the moments computation requires a segmentation step whose efficiency is dependent on the organ shape and which is not robust to organ topology changes.

To overcome the image processing limitation, recent works have proposed to use directly as visual features the intensity value of the image pixels. In [11], an approach based on the speckle correlation observed in successive US images is detailed. However, a step of learning of the decorrelation curves is required to control the out-of-plane motions of the probe and this local approach is only fitted for tracking tasks. In [12], we proposed an intensity-based approach to control the 6 dof of a 2D probe, where the interaction matrix relies on the estimation of the 3D image gradient. In this previous work, this estimation was made possible once at the initial pose of the probe, which restrained the method to tracking applications. In this paper we present a new way to estimate on-line the 3D image gradient, which allows the computation of the current interaction matrix of the system and the extension of the intensity-based approach to positioning tasks. The medical applications we consider are needle insertion procedures or medical diagnoses where the visual servoing can be used to reach and stabilize the view of an organ or tumor while compensating for physiological

motions of the patient. The second contribution of this work is the modelling of the US intensity-based approach for dealing with a bi-plane probe.

The structure of our paper is as follows. We initially detail the intensity-based approach with the on-line estimation of the interaction matrix. Validation and improvements with respect to the use of the initial interaction matrix for positioning tasks are demonstrated in simulation. In Section III, we propose to extend this intensity-based approach to take into account the visual information extracted from both orthogonal images provided by an US bi-plane device. Experiments of tracking and positioning tasks with 2D and bi-plane US sensors are then presented in Section IV. Finally, perspectives and concluding remarks are given in Section V.

## II. INTENSITY-BASED CONTROL

An image-based visual servoing control scheme consists in minimizing the error between a current set of visual features  $\mathbf{s}$  and a desired one  $\mathbf{s}^*$ . In the case of a number  $p$  of visual features exceeding the number  $m$  of dof of the controlled system, a combination matrix  $\mathbf{C}$  of size  $m \times p$  and full rank  $m$  is introduced to define the task function to minimize:

$$\mathbf{e}(t) = \mathbf{C}(\mathbf{s}(t) - \mathbf{s}^*). \quad (1)$$

In an eye-in-hand system, the instantaneous velocity applied to the imaging sensor  $\mathbf{v}_p$  is computed from the task function and the interaction matrix  $\mathbf{L}_s$ . More particularly, the classical visual servoing control law described in [1] is based on an exponential decrease of the task function ( $\dot{\mathbf{e}} = -\lambda \mathbf{e}$ ,  $\lambda > 0$ ) and presents the best behavior with  $\mathbf{C} = \widehat{\mathbf{L}}_s^+$ , which yields to:

$$\mathbf{v}_p = -\lambda \left( \widehat{\mathbf{L}}_s^+ \widehat{\mathbf{L}}_s \right)^{-1} \left( \widehat{\mathbf{L}}_s^+ (\mathbf{s}(t) - \mathbf{s}^*) \right) = -\lambda \widehat{\mathbf{L}}_s^+ (\mathbf{s}(t) - \mathbf{s}^*) \quad (2)$$

### A. Intensity features

With visual servoing techniques, the efficiency of the control law is highly dependent of the choice of the visual features. Typically, geometric features, which can be extracted by various methods based on image similarity measure, contour segmentation or image processing (threshold and morphological operators), are well-adapted to control the different motions of a robotic system. However, the processing of the US images is often more complex than with other imaging modalities due to the noise, called speckle, generated by the propagation of the US waves in the soft tissues. In this paper, we propose to avoid any US image processing step by choosing as visual feature the image itself. In this case, the visual features vector  $\mathbf{s}$  corresponds to the intensity values of the pixels contained in a region of interest (ROI) of the US image:

$$\mathbf{s} = \{I(1,1), \dots, I(u,v), \dots, I(M,N)\}, \quad (3)$$

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 coordinates  $(u,v)$  in the US image. We detail in the following the computation of the interaction matrix  $\mathbf{L}_s$  that

links the variation of these image features to in-plane and out-of-plane motions of the US probe in order to control the six dof of the device.

### B. Modelling of the interaction

Given  $\mathcal{R}_p(\mathbf{x}_p, \mathbf{y}_p, \mathbf{z}_p)$  the frame attached to the US probe where  $(\mathbf{x}_p, \mathbf{y}_p)$  defines the image plane and  $\mathbf{z}_p$  corresponds to the elevation axis, the coordinates  $\mathbf{x} = (x, y, z)$  of an image point in this frame are such as:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} s_x(u - u_0) \\ s_y(v - v_0) \\ 0 \end{pmatrix}, \quad (4)$$

with  $(s_x, s_y)$  the image pixel size and  $(u_0, v_0)$  the pixel coordinates of the image center, and where  $z = 0$  since the considered visual features belong to the US image plane.

The interaction matrix estimation relies on the constancy of the US wave reflection by a physical 3D point. Given such a 3D point at the position  $\mathbf{x}$  at the time  $t$  that moves to the new position  $\mathbf{x} + d\mathbf{x}$  at the time  $t + dt$  and considering that the US reflection is converted in an intensity value in a B-mode US image, the US reflection conservation yields to the following intensity conservation equation:

$$I(\mathbf{x} + d\mathbf{x}, t + dt) - I(\mathbf{x}, t) = 0, \quad (5)$$

that can be expanded in the form of a first order Taylor series:

$$\frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial z} dz + \frac{\partial I}{\partial t} dt = 0. \quad (6)$$

Then, the time variation of each pixel intensity  $I(u, v)$  can be expressed as a function of the corresponding 3D point motion :

$$\dot{I}(u, v) = -\nabla I_{(u,v)} \dot{\mathbf{x}}, \quad (7)$$

with  $\nabla I_{(u,v)} = (\nabla I_x \nabla I_y \nabla I_z)$  the 3D image gradient associated to the pixel  $(u, v)$ .

According to the kinematics fundamental relationship, the velocity of the 3D point  $\dot{\mathbf{x}} = (\dot{x}, \dot{y}, \dot{z})$  can be expressed as a function of the in-plane and out-of-plane components of the probe motion  $\mathbf{v}_p$ , through the interaction matrix  $\mathbf{L}_x$ :

$$\dot{\mathbf{x}} = \mathbf{L}_x \mathbf{v}_p, \quad \mathbf{L}_x = \begin{bmatrix} -1 & 0 & 0 & 0 & -z & y \\ 0 & -1 & 0 & z & 0 & -x \\ 0 & 0 & -1 & -y & x & 0 \end{bmatrix}. \quad (8)$$

From (7) and (8), the interaction matrix  $L_{I(u,v)}$  of size  $1 \times 6$  associated to the visual feature  $I(u, v)$  is written as:

$$L_{I(u,v)} = [-\nabla I_x \quad -\nabla I_y \quad -\nabla I_z \quad -y \nabla I_z \quad x \nabla I_z \quad y \nabla I_x - x \nabla I_y], \quad (9)$$

and the complete interaction matrix  $\mathbf{L}_s$ , defined as  $\dot{\mathbf{s}} = \mathbf{L}_s \mathbf{v}_p$ , is built by stacking the  $M \times N$  matrices  $L_{I(u,v)}$ .

### C. On-line gradient estimation

To control the six dof of the US probe, the variation of the visual features is related to the in-plane and out-of-plane motions of the sensor through the interaction matrix. More particularly the out-of-plane component of the 3D image gradient  $\nabla I_z$  is essential to compute this interaction matrix.

In [12], we used additional parallel images to compute the 3D gradient with a 3D filter applied on this set of images. However with a conventional 2D US probe mounted on a robotic arm, a small back and forth translational motion along the elevation direction was required to acquire these images. In this case, the interaction matrix was estimated once at the desired pose of the probe and used without being updated to perform a tracking task.

However, for positioning tasks the interaction matrix has to be estimated on-line, and we propose here a new approach to estimate the gradient component  $\nabla I_z$  from the past images acquired during the positioning task. This estimation is based on the Taylor expansion of the pixel intensity:

$$I(\mathbf{x} + \mathbf{dx}) = I(\mathbf{x}) + \nabla I_x dx + \nabla I_y dy + \nabla I_z dz. \quad (10)$$

This equation is rewritten as:

$$Y = \Phi \theta, \quad (11)$$

with:

$$\begin{cases} Y &= I(\mathbf{x} + \mathbf{dx}) - (I(\mathbf{x}) + \nabla I_x dx + \nabla I_y dy) \\ \Phi &= dz \\ \theta &= \nabla I_z \end{cases}, \quad (12)$$

where  $\theta$  is the parameter of the system to estimate and where  $Y$  and  $\Phi$  can be measured at each iteration. The pixel motion  $\mathbf{dx} = (dx, dy, dz)$  due to the probe motion is computed with the robot odometry while the intensity values of this pixel and the in-plane gradient components ( $\nabla I_x, \nabla I_y$ ) can be directly measured in the image.

To take into account the past values of the measures and then increase the accuracy of this estimation, we propose to solve the equation (11) by using the recursive least squares (RLS) algorithm. The RLS algorithm consists in computing the estimate of the parameter  $\hat{\theta}_{[k]}$  that minimizes the cost function  $J(\hat{\theta}_{[k]})$ , where  $J(\hat{\theta}_{[k]})$  corresponds to the squared sum of the prediction errors  $E_{[k]} = Y_{[k]} - \Phi_{[k]} \hat{\theta}_{[k]}$ , weighted by a forgetting factor  $0 < \beta \leq 1$  used to lower the influence of past data:

$$J(\hat{\theta}_{[k]}) = \sum_{j=k_0}^k \beta^{k-j} (Y_{[j]} - \Phi_{[j]} \hat{\theta}_{[j]})^2. \quad (13)$$

The parameter  $\hat{\theta}_{[k]}$  that minimizes this convex function sets to zero its derivative and is recursively defined as follows [13]:

$$\hat{\theta}_{[k]} = \hat{\theta}_{[k-1]} + \delta(e(k)) F_{[k]} \Phi_{[k]} (Y_{[k]} - \Phi_{[k]} \hat{\theta}_{[k-1]}), \quad (14)$$

where the factor  $F$  is also recursively defined as:

$$F_{[k]}^{-1} = \beta F_{[k-1]}^{-1} + \delta(e(k)) \Phi_{[k]} \Phi_{[k]} + \epsilon_0. \quad (15)$$

The term  $\epsilon_0$  is known as a stabilization term and ensures that the factor  $F_{[k]}$  remains invertible even if the signal  $\Phi_{[k]}$  is not sufficiently excited, which corresponds to a pure in-plane motion of the probe. Moreover, a dead zone is added to avoid the update of the parameter estimation when the euclidean norm of the prediction error  $e(k) = \|E_{[k]}\|$  is low.

The function  $\delta(e(k))$  that characterizes the dead zone is chosen as defined in [14]:

$$\delta(e(k)) = \begin{cases} 1 - \frac{\delta_0}{e(k)} & \text{if } e(k) \geq \delta_0 \\ 0 & \text{else} \end{cases}, \quad (16)$$

where  $\delta_0$  is an estimation of the maximum value of the prediction error due to the noise and disturbance.

#### D. Simulation validation

To validate our approach, we use a software simulator that we have developed to reconstruct and display a dense volume from a set of parallel images. In addition to this display functionality, the simulator allows the control of a virtual 2D probe that generates an US image by cubic interpolation process. An US complete volume of the right kidney of a realistic abdominal phantom is loaded in the simulator. This volume is created from a set of 335 parallel images of size  $250 \times 250$ , the voxel size is  $0.6 \times 0.6 \times 0.3 \text{ mm}^3$ .

We 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 probe on the kidney volume and we consider the corresponding US scan as the desired one. Then the probe is moved away to a new pose where the observed organ section is considered as the initial image. The results of the positioning task are given in Fig. 1. A visual error is defined as the euclidean norm of the error between the current and desired visual features vectors  $\mathcal{C} = (\mathbf{s} - \mathbf{s}^*)^\top (\mathbf{s} - \mathbf{s}^*)$ . The difference images between the current and the desired US scans are displayed for the initial and final probe poses. The uniform gray color of this difference image after the convergence of the algorithm shows the success of the positioning task since the final image perfectly coincides with the desired one.

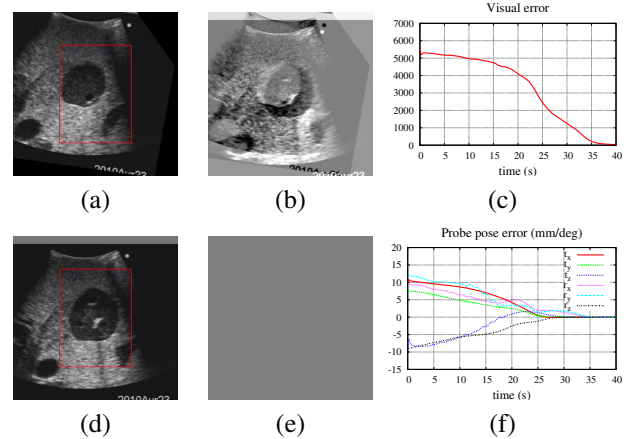


Fig. 1. Positioning task using visual features extracted from a single US image. The visual convergence from the initial US view (a) to the final one (d) is shown by the corresponding difference image with the desired view (respectively (b) and (e)). The visual (c) and pose (f) convergence is displayed during the positioning task.

During the five first iterations of the algorithm an open-loop out-of-plane translation is applied to the probe in order to initialize the estimation of the gradient component  $\nabla I_z$ .

Then, during the convergence process, a new estimate of this parameter is computed each time enough motion along the  $z$ -axis is observed. The evolution of the value of this estimate for one pixel is shown in Fig. 2.

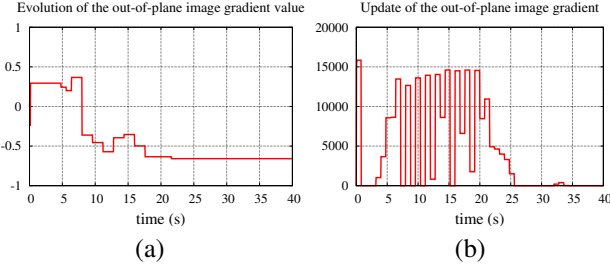


Fig. 2. Estimation of the  $\nabla I_z$  parameter. (a) Evolution of this value for one pixel. (b) Number of parameters updated throughout the positioning task.

These results can be compared to the ones obtained with no update of the interaction matrix. We consider the same initial and desired poses of the probe and the 3D image gradient is computed from a set of five parallel images acquired at the initial probe pose. The interaction matrix is then estimated and used without being updated during the convergence process. The Fig. 3 shows the behavior of the control law, which fails to converge to the desired pose.

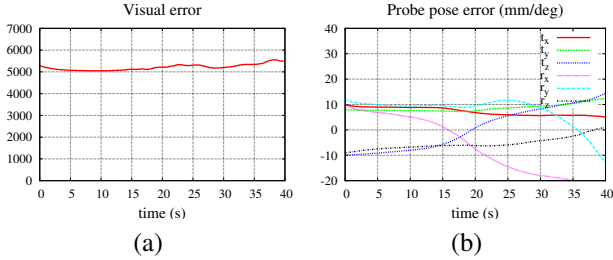


Fig. 3. Failure of the positioning task using the initial interaction matrix.

### III. BI-PLANE APPROACH

Among the new generation of US transducers, the XYPA2.5 device commercialized by the Vernon company is a 2.5 MHz bi-plane US probe that provides two orthogonal images at a frame rate of 20Hz (see Fig. 4). With such a frame rate, this sensor can therefore be used for image-guided control of a robotic system, where the additional information provided by the orthogonal image plane can be integrated to improve the performance of the control. The second contribution of this work is then the modelling of the interaction to take into account the visual information extracted from both orthogonal images of a bi-plane probe.

#### A. Modelling of the interaction

In each image plane  $US_i$ , a set of image pixels  $s_i$  is extracted from the ROI and the new features vector is  $\mathbf{s} = (s_0, s_1)$ . The variation of these image features  $\dot{\mathbf{s}}$  is related to the instantaneous velocity  $\mathbf{v}_{US_i}$  of the corresponding image plane through the interaction matrix previously described:

$$\begin{cases} \dot{s}_0 &= \mathbf{L}_{s_0} \mathbf{v}_{US_0} \\ \dot{s}_1 &= \mathbf{L}_{s_1} \mathbf{v}_{US_1} \end{cases} \quad (17)$$

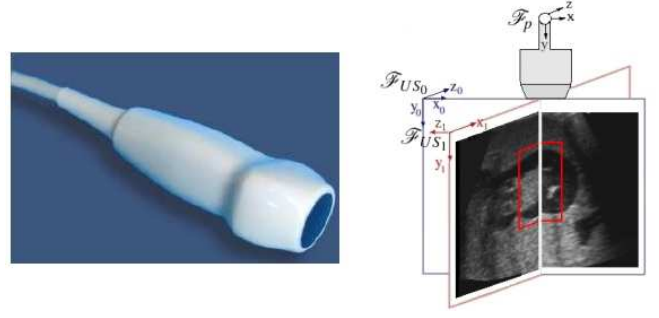


Fig. 4. The XYPA2.5 bi-plane device (Vernon) (left) provides two orthogonal US images. The probe frame  $\mathcal{F}_p$  is aligned with the frame of the image plane  $US_0$  and the plane  $US_1$  corresponds to a rotation of  $90^\circ$  around the  $y$ -axis (right).

Both image planes being rigidly attached to the probe, their velocities  $\mathbf{v}_{US_i}$  can be expressed in function of the instantaneous velocity of the probe  $\mathbf{v}_p$ :

$$\forall i \in \{0, 1\}, \mathbf{v}_{US_i} = {}^{US_i} \mathbf{W}_p \mathbf{v}_p \quad (18)$$

with:

$${}^{US_i} \mathbf{W}_p = \begin{pmatrix} {}^i \mathbf{R}_p & [{}^i \mathbf{t}_p]_{\times} {}^i \mathbf{R}_p \\ 0_3 & {}^i \mathbf{R}_p \end{pmatrix}, \quad (19)$$

where  ${}^i \mathbf{t}_p$  and  ${}^i \mathbf{R}_p$  are the translation vector and the rotation matrix of the probe frame  $\mathcal{F}_p$  expressed in the image frame  $\mathcal{F}_{US_i}$ . For the control of the probe, we consider that  $\mathcal{F}_{US_0}$  coincides with  $\mathcal{F}_p$ , hence:

$$\begin{cases} {}^{US_0} \mathbf{W}_p = \mathbf{I}_6 \\ {}^{US_1} \mathbf{W}_p = \begin{pmatrix} \mathbf{R}_y & 0_3 \\ 0_3 & \mathbf{R}_y \end{pmatrix} \end{cases} \quad \mathbf{R}_y = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}. \quad (20)$$

The interaction matrix associated to the features vector  $\mathbf{s}$  extracted from both image planes of the bi-plane probe is then:

$$\mathbf{L}_s = \begin{bmatrix} [-\nabla I_x & -\nabla I_y & -\nabla I_z & -y\nabla I_z & x\nabla I_z & y\nabla I_x - x\nabla I_y]_{\forall (u,v) \in US_0} \\ [\nabla I_z & -\nabla I_y & -\nabla I_x & -y\nabla I_x + x\nabla I_y & x\nabla I_z & -y\nabla I_z]_{\forall (u,v) \in US_1} \end{bmatrix} \quad (21)$$

#### B. Simulation validation and discussion

For this simulation the same desired pose of the probe as for the 2D approach is considered and the desired features are extracted from both orthogonal images of a virtual bi-plane probe. A different initial pose is then chosen, farther than in the previous simulation and from which the one-plane algorithm falls in a local minimum. The results of the bi-plane positioning task are given in Fig. 5. In this case, no open-loop out-of-plane translation is required to initialize the estimation of the components  $\nabla I_z$  associated to each pixel of both images. At the first iteration of the algorithm, these out-of-plane gradient components are set to zero, as is shown on the probe pose evolution (see Fig. 5(a)) by the null initial rotational velocity of the probe around its  $y$ -axis. The five other dof of the probe are controlled from the in-plane gradient components, generating thus an out-of-plane motion of the probe that allows the estimation of

the parameters  $\nabla I_z$  with the method proposed in Section II-C. During the positioning process, these parameters are then updated each time enough out-of-plane motion is applied to the corresponding pixel until convergence.

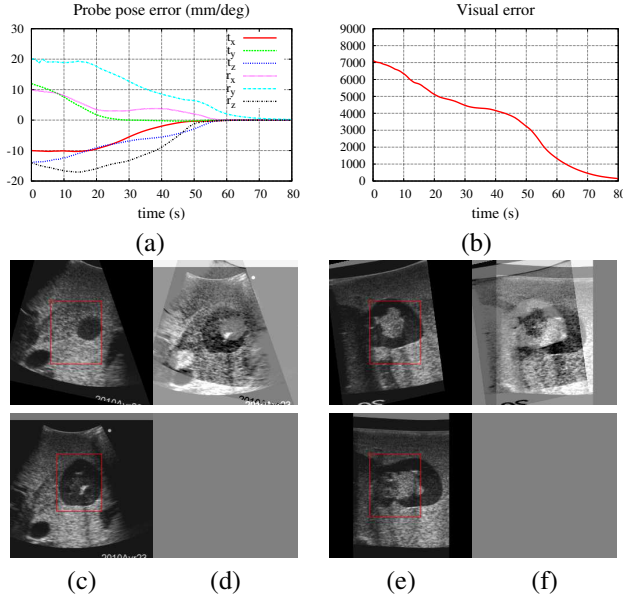


Fig. 5. Positioning task with a bi-plane US probe. The convergence is validated in term of pose (a) and visual (b) errors. (c) and (e) are both images of the bi-plane probe at its initial (top) and final (bottom) pose. (d) and (f) are the corresponding difference images with the desired US image.

The contribution of the gradient component  $\nabla I_x$  computed from the image plane  $US_1$  in the out-of-plane motion of the US probe allows the on-line estimation of the out-of-plane gradient without initialization step. On the other hand, the extraction of visual features from two orthogonal images increases the convergence domain of the visual servoing.

#### IV. EXPERIMENTAL RESULTS

With the intensity-based approach we propose to provide assistance to a surgeon by reaching and following a desired organ cross section and thus stabilizing the view of the US probe. In this work, we validate both positioning and tracking tasks. First we implement the on-line estimation of the 3D image gradient on a convex 2-5 MHz 2D US probe to perform a tracking task while the phantom undergoes fast motions. Then, in a second part, we consider a bi-plane 2-5 MHz US probe to perform local positioning tasks.

##### A. Experimental setup

Experiments are performed with an anthropomorphic robotic arm equipped with an US transducer and a force sensor (see Fig. 6). We combine the visual control with a force control that guarantees a constant force of 1N applied to the phantom and servo the translational motion along the y-axis of the probe frame. As we chose 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 y-translational velocity component due to the image control is not applied to the probe. To deal with this issue, we propose

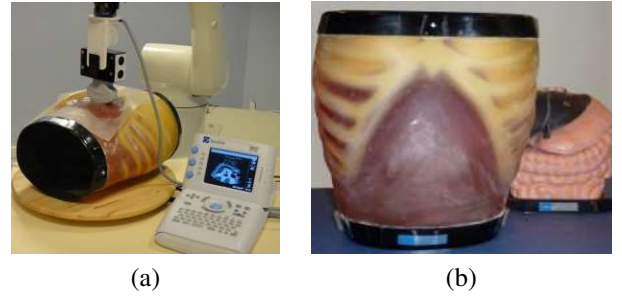


Fig. 6. Experimental setup. A six dof anthropomorphic robotic system (a) is equipped with a force sensor and a bi-plane or 2D US probe which provides images of a realistic abdominal phantom (b).

to separate the dof controlled by the force control from the others. The five velocity components corresponding to the translations along the  $x$  and  $z$  axis of the probe and to the three rotations are applied physically to the US device while the last component corresponding to the translation along the  $y$  axis is virtually applied to a window containing the ROI and included in the US image.

##### B. Tracking with a 2D probe

We position the 2D US probe on the abdominal phantom and we define a ROI in the observed US B-scan that contributes to the visual control (see Fig. 7(a)). The force/vision servo process 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 these disturbances 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. 7 and in the attached video.

The interaction matrix, initialized at the desired pose, is updated during the tracking task from the image gradient extracted directly from the current image (for the 2D components  $\nabla I_x$  and  $\nabla I_y$ ) and estimated from the image measures as described in Section II-C (for its out-of-plane component  $\nabla I_z$ ). Despite the important disturbances applied to the phantom, which generate some delays in the tracking, the US probe converges to the desired image.

##### C. Positioning task with a bi-plane probe

In this second experiment, the bi-plane probe is mounted on the robotic arm and positioned on the abdominal phantom to acquire both desired images. From a different initial pose of the probe, the bi-plane intensity-based approach is then used to reach the desired images. Contrarily to experiments performed with a 2D probe, the image gradient is not initialized at the desired probe pose. The in-plane components are computed directly from the initial images and the out-of-plane component is set to zero at this pose, before being estimated when the probe moves. The results of the positioning task are given in Fig. 8.



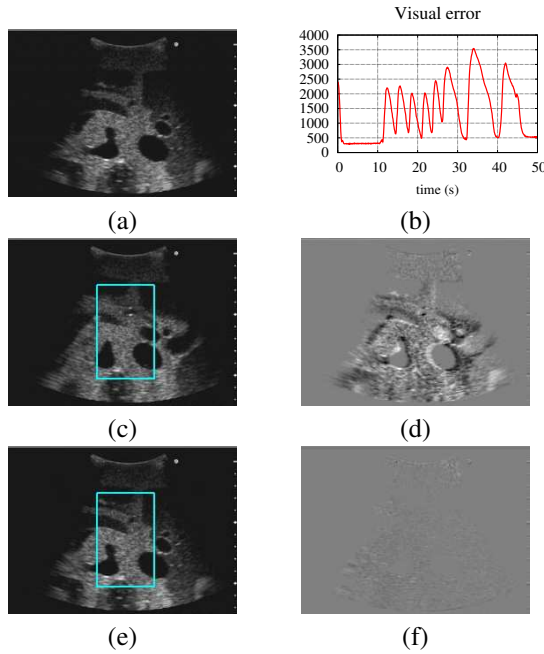


Fig. 7. Tracking of an abdominal cross-section (a) with a 2D probe. The tracking delay of the 2D probe is observed through the evolution of the visual error (b) and through the US image (c) and difference image (d) corresponding to the maximum visual error (at  $t = 33s$ ). Nevertheless, at the end of the disturbance, the probe reaches its desired pose as is shown by the final US image (e) and difference image (f).

## V. CONCLUSION AND PERSPECTIVES

This paper presented new methods to improve the accuracy of the US intensity-based approach by computing the current interaction matrix in the control law and by considering a bi-plane device. We have implemented and validated these methods in simulation environment but also on a robotic system with two different US sensors. In further works, we will propose a qualitative and quantitative comparison of the 2D and bi-plane approaches. Finally we plan to propose a full autonomous process of robotic assistance for a medical gesture, where both in-plane and out-of-plane motions of an US probe are controlled to reach and follow one or several desired views of an organ while compensating for physiological motions of a patient. This way, all along the medical gesture, the surgeon could have a display of a stabilized view of a chosen organ or tumor and could automatically navigate from one view to another.

### ACKNOWLEDGMENT

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

### REFERENCES

- [1] B. Espiau, F. Chaumette and P. Rives, A new approach to visual servoing in robotics. *In IEEE Trans. on Robotics*, 8(3):313-326, 1992.
- [2] D. Uecker, L. Cheolwhan, Y. Wang, Y. Wang, Automated instrument tracking in robotically assisted laparoscopic surgery. *In Journal of Image Guided Surgery*, vol. 1, pp. 308-325, 1995.
- [3] P. Abolmaesumi, S. Salcudean, W. Zhu, M. Sirouspour, and S. DiMaio, Image-guided control of a robot for medical ultrasound. *In IEEE Trans. on Robotics*, vol. 18, no. 1, February 2002.

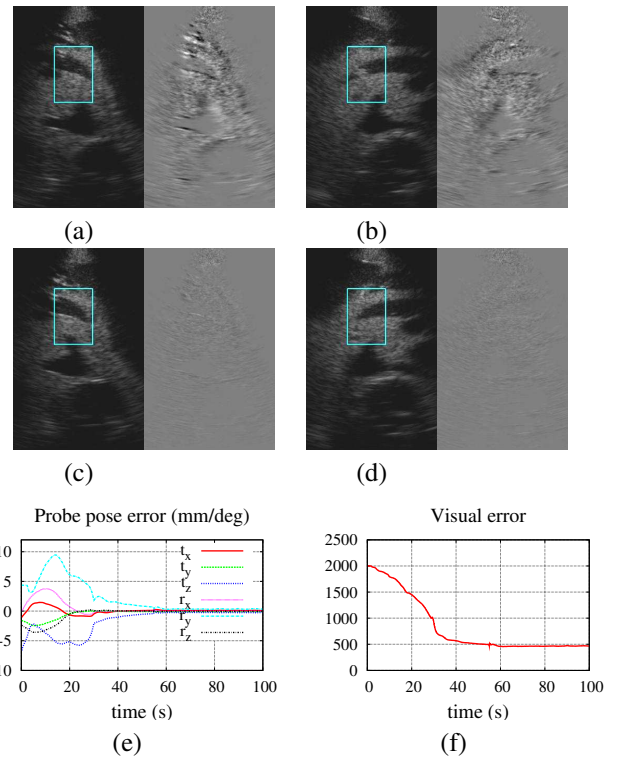


Fig. 8. Positioning task with the bi-plane probe. The visual convergence (f) from the initial US views (a) and (b) to the final ones (c) and (d) is shown by the corresponding images difference with the desired views. With no motion applied to the phantom, the convergence in term of pose (e) is computed with the robot odometry.

- [4] J. Hong, T. Dohi, M. Hashizume, K. Konishi, N. Hata, A motion adaptable needle placement instrument based on tumor specific ultrasonic image segmentation. *In 5th Int. Conf. on Med. Image Computing and Computer Assisted Intervention, MICCAI'02*, pp122-129, Tokyo, Japan, 2002.
- [5] P.M. Novotny, J.A. Stoll, P.E. Dupont and R.D. Howe, Real-time visual servoing of a robot using three-dimensional ultrasound. *In IEEE Int. Conf. on Rob. and Auto., ICRA'07*, pp2655-2660, Roma, Italy, 2007.
- [6] D. Lee, N. Koizumi, K. Ota, S. Yoshizawa, A. Ito, Y. Kaneko, Y. Matsumoto, and M. Mitsuishi, Ultrasound-based visual servoing system for lithotripsy. *In IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'07*, pp.877 -882, 2007.
- [7] M.A. Vitrani, H. Mitterhofer, N. Bonnet, G. Morel, Robust ultrasound-based visual servoing for beating heart intracardiac surgery. *In IEEE Int. Conf. on Rob. and Auto., ICRA'07*, pp3021-3027, Roma, Italy, 2007.
- [8] M. Sauvee, P. Poignet, E. Dombre, US image based visual servoing of a surgical instrument through non-linear model predictive control, *In Int. Journal of Robotics Research*, vol. 27, no. 1, January 2008.
- [9] R. Mebarki, A. Krupa and F. Chaumette, 2D ultrasound probe complete guidance by visual servoing using image moments. *In IEEE Trans. on Robotics*, vol. 26, no. 2, pp 296-306; 2010.
- [10] C. Nadeau, A. Krupa, A multi-plane approach for ultrasound visual servoing : application to a registration task. *In IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'10*, Taipei, 2010.
- [11] A. Krupa, G. Fichtinger, G. Hager, Real time motion stabilization with B-mode ultrasound using image speckle information and visual servoing. *In Int. Journal of Robotics Research, IJRR*, 2009.
- [12] C. Nadeau, A. Krupa, Intensity-based direct visual servoing of an ultrasound probe. *IEEE International Conference on Robotics and Automation, ICRA'11*, Shanghai, 2011.
- [13] G. Kreisselmeier, Stabilized least-squares type adaptive identifiers. *In IEEE Trans. on Automatic Control*, vol. 35, no.3, pp.306310, 1990.
- [14] R.H. Middleton, G.C. Goodwin, D.J. Hill, and D.Q. Mayne, Design issues in adaptive control. *In IEEE Trans. on Automatic Control*, vol. 33, pp.5058, 1988.