Moments-Based Ultrasound Visual Servoing: From Mono to Multi-plane Approach

Caroline Nadeau, Alexandre Krupa, Jan Petr, Christian Barillot

Moments-based ultrasound visual servoing: from mono to multi-plane approach

Caroline Nadeau[1,2], Alexandre Krupa[1], *Member, IEEE,* Jan Petr[1,4] and Christian Barillot[1,3], *Senior Member, IEEE*

[1]INRIA Rennes-Bretagne Atlantique and IRISA (CNRS UMR 6074), Rennes, France
[2]University of Rennes I, Rennes, France
[3]INSERM U746, Rennes, France
[4]Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany

***Abstract*—This paper presents a new image-based visual servoing approach to control a robotic system equipped with an ultrasound imaging device. The presented method allows an automatic positioning of the probe with respect to an object of interest. Moments-based image features are computed from three orthogonal ultrasound images to servo in-plane and out-of-plane motions of the system. An efficient segmentation method, based on graph cut strategy, is proposed to extract the object contour in each image plane. Simulation results demonstrate that this approach improves upon techniques based on a single 2D US image in terms of probe positioning. Our method was also validated from robotic experiments performed on an ultrasound phantom with the use of a motorized 3D probe that provides the three US images.**

***Index Terms*—Visual servoing, ultrasound images, graph cut segmentation, moment features.**

I. Introduction

By definition, the term visual servoing designates the control of the motion of a dynamic system using a vision sensor. The variation of the visual data provided by this sensor is linked to its motion with respect to the scene by a matrix called interaction matrix [1]. An estimation or the analytic form of this matrix is then used in a closed-loop control scheme in order to move the sensor so as to minimize the error between the current visual information and the desired one. Visual servoing has been mainly used with perspective camera but this formalism remains valid in the case of other vision sensors such as the ultrasound (US) sensor, which is here assimilated to a visual sensor since it provides gray scale 2D images in B-scan mode.

The first US visual servoing has been proposed by Abolmaesumi *et al.* [2] to control the in-plane motions of a 2D US probe while the out-of-plane motions of the probe are tele-operated. To servo these three degrees of freedom (DOF), the 2D coordinates of two arteries centers are used as visual features. The detection of these image points requires a preliminary step of segmentation of the artery contour. Five extraction methods are compared, which are based on image similarity measure such as cross correlation and sequential similarity detection or on contour segmentation by a Star [3] or Snake algorithm. In addition to this "eye-in-hand" configuration where the control is directly applied to the US probe manipulated by a robotic system, the US visual servoing can also be used to control a medical tool under US guidance in an "eye-to-hand" configuration. In [4], two in-plane DOF of a needle-insertion robot are then controlled by visual servoing to perform a percutaneous cholecystostomy while compensating for involuntary patient motions. The axis of the needle, rigidly aligned within the US probe plane is extracted with the Hough transform and the target tumor is detected using an active contour. In the same way, in [5], a cross-shaped pattern is used to represent an anatomic target and a passive marker is fixed to the tool. A Radon transform is then performed to extract these features in a 3D US image. The tool detection is well performed but Hough and Radon transforms are specific for identifying long axes or detecting intersecting lines and can not be extended to detect anatomic targets. For a lithotripsy procedure [6], which consists in the removal of kidney stones using high-intensity focused ultrasound (HIFU), two US probes and the HIFU transducer are mounted on the end effector of a XYZ stage robot to follow a target kidney stone while compensating for 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.

Some authors have proposed solutions to control the out-of-plane motions of a US probe or a surgical tool by visual servoing. In [7], a robotic system is proposed to track a surgical instrument and move it to a desired target. 3D US images provided in real time by a matrix-array 3D probe are processed to localize respective positions of the target and the instrument tip, then the position error is used to control four DOF of the robotized tool. However matrix-array 3D probes provide small and low-quality volumes which limit the amplitude of the surgical tool. With a 2D probe, Vitrani *et al.* have developed solutions to servo the four non-constrained DOF of a forceps inserted through a trocar in order to reach a desired pose. From the current and desired images of the probe, a visual servoing loop is implemented to move the tool while maintaining it in the US image plane. To control these four DOF, the coordinates of two image points corresponding to the intersection of the forceps jaws with the image plane are used as visual features [8], [9]. More recently, Nakadate *et al.* described in [10] an 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. A previous step of acquisition of several parallel images around the target image is required and the approach is applied to a tracking task.

Finally, few approaches have been proposed to control the six DOF of the probe. In [11], 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. However, a region of fully developed speckle has to be segmented and a step of learning of the speckle decorrelation curves is required. In [12], Nadeau and Krupa considered intensity features to control the six DOF of a 2D probe, using the 3D image gradient to express the variation of the pixel intensities to the probe motion. However, due to the local nature of the considered features, this intensity-based approach is more particularly dedicated for tracking tasks and often leads to local minima for positioning tasks from a remote initial pose. On the contrary, geometric features are well-adapted to positioning tasks. In [13], the six DOF of a US probe are controlled with a moments-based approach where the six visual features are computed from 2D moments of the object cross-section in one US image. The analytic expression of the corresponding interaction matrix is modeled and the visual servoing is implemented to perform positioning tasks. The obtained results show a good behavior of the approach in terms of minimization of the visual error but only a local convergence of the probe is guaranteed. In particular, in the case of a rough symmetric anatomic target, two different cross sections with similar geometric properties can be observed.

Further to the aforementioned work, we present here a new set of visual features that allows a global convergence of the control law even when considering rough symmetric objects. Six geometric features are computed from the 2D moments of the object cross-section, segmented in several orthogonal image planes, to control the six DOF of the US probe. Moreover, to guarantee a more robust segmentation of the object, we develop a graph cut strategy instead

This work was presented in part at the IEEE Int. Conf. on Intelligent Robots and Systems, Taipei, Taiwan, October 2010 [14].
Contact: Alexandre.Krupa@inria.fr

of the active contours considered in [13].

The structure of our paper is as follows. We firstly present the moments-based visual servoing approach with the computation of the image 2D moments from the object segmented in the image. The mono-plane strategy [13] is then recalled and we show as a first contribution its limitations through simulation results in Section II-A. In Section II-B, the second contribution of this work, which is the multi-plane approach, is described and validated in simulation environment. Note that we derived this approach from our preliminary work [14] that presented an offline multimodal image registration method whereas in this paper we address the control of a real robotic system actuating an 3D ultrasound probe. We present in Section III the third contribution of the work, which is a real-time US segmentation method based on a graph-cut algorithm. This robust segmentation is then used with the multi-plane visual servoing to perform positioning tasks using a robotic arm manipulating a 3D motorized probe that interacts with an ultrasound phantom. These robotic results are gathered in Section IV and allow us to conclude on the benefits of our approach.

II. Moments-based ultrasound 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 $\mathbf{e}(t) = \mathbf{s}(t) - \mathbf{s}^*$ with:

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

where λ is a positive gain tuning the decrease time of the visual error. In an eye-in-hand configuration, $\mathbf{v}_c$ 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 $\mathbf{v}_c$ ($\dot{\mathbf{s}} = \mathbf{L}_\mathbf{s} \mathbf{v}_c$).

The choice of suitable visual features is crucial to ensure a good behavior of the control law. In this work, we are interested in image moments, which have been first used in camera-based visual servoing [15]. They have further been introduced for US visual servoing by Mebarki *et al.* [13] to control the six DOF of a 2D US probe. These geometric features show good properties for US image-based control since they are robust to image noise and since the low order moments characterize the geometry of the object of interest in the image. In our case, the image moments m_{ij} of order $i+j$ are computed after the extraction of the contour C of the object of interest in the considered image that we perform thanks to the segmentation method proposed in Section III:

$$m_{ij} \quad = \quad \frac{-1}{j+1} \oint_C x^i \, y^{j+1} \, dx \tag{2}$$

A. Existing mono-plane approach and discussion

Six geometric features are proposed in [13] to define the features vector $\mathbf{s}$. They represent the section of the object in the US plane by its mass center coordinates (x_g, y_g) and its main orientation angle α in the image which are representative of the in-plane motions of the probe and present good decoupling properties. The area a of the object section, invariant to in-plane motions, and ϕ_1 and ϕ_2 depending respectively of moments of order 2 and 3 which are invariant to the image scale, translation, and rotation are chosen to control out-of-plane motions. These features are computed from the image moments as follows (see [13] for details of the analytical calculus):

$$\left\{ \begin{array}{lcl} x_g & = & m_{10}/m_{00} \\ y_g & = & m_{01}/m_{00} \\ \alpha & = & \frac{1}{2} \, arctan(\frac{2\,\mu_{11}}{\mu_{20}-\mu_{02}}) \\ a & = & m_{00} \\ \phi_1 & = & \frac{{\mu_{11}}^2 - \mu_{20}\mu_{02}}{(\mu_{20}-\mu_{02})^2 + 4{\mu_{11}}^2} \\ \phi_2 & = & \frac{(\mu_{30}-3\mu_{12})^2 + (3\mu_{21}-\mu_{03})^2}{(\mu_{30}+\mu_{12})^2 + (\mu_{21}+\mu_{03})^2} \end{array} \right. \tag{3}$$

The computation of the interaction matrix used to control in-plane and out-of-plane motions of the US probe is based on the time variation of moments of order $i+j$ expressed as a function of the probe velocity: $\dot{m}_{ij} = \mathbf{L}_{\mathbf{m_{ij}}} \mathbf{v}_c$ with $\mathbf{L}_{\mathbf{m_{ij}}} = [m_{v_x} \; m_{v_y} \; m_{v_z} \; m_{\omega_x} \; m_{\omega_y} \; m_{\omega_z}]$. The components $(m_{v_x}, m_{v_y}, m_{\omega_z})$ related to the in-plane probe velocity are directly expressed from image moments. However the remaining components $(m_{v_z}, m_{\omega_x}, m_{\omega_y})$ also depend on the 3D normal vector to the object surface which has to be estimated in each contour point. The final form of the resulting interaction matrix, whose detailed form is given in [13], can be rewritten as:

$$\mathbf{L}_\mathbf{s} = [\mathbf{L}_{x_g} \, \mathbf{L}_{y_g} \, \mathbf{L}_\alpha \, \mathbf{L}_a \, \mathbf{L}_{\phi_1} \, \mathbf{L}_{\phi_2}]^T \tag{4}$$

We propose to analyze the behavior of the mono-plane moments-based visual servoing without considering errors due to the segmentation process or the normal estimation algorithm. For this purpose we use a geometrical simulator that mathematically generates the intersection of the plane of a virtual probe with a volume constituted of four spheres of different radii. Given a pose of the virtual probe, a binary image of the object hull is created and its contour is extracted thanks to the use of a basic connected-component detection algorithm. Moreover, considering the particular geometry of the object, the normal vector in each point of its surface is perfectly known. The results of a positioning task obtained by applying the visual control law (1) with the selected visual features (3) are presented in Fig. 1. The virtual probe is positioned to a desired pose and the corresponding desired visual features vector is saved. A different pose is then taken as initial probe pose and the visual servoing is launched. The binary images (a) and (b) are respectively the initial and final image of the probe, where the desired object cross-section is delineated in red. The initial pose error is:

$$\Delta\mathbf{r}_{init}(mm, deg) = [-7, -14, 5, -8, -12, 12].$$

The three first components of this vector describe the error in translation and the three last the error in rotation (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 choice of the six visual features (3) extracted from one image plane ensures at least the achievement of the positioning task in terms of visual error. However, the information belonging to one single plane is not always sufficient to characterize the probe pose. Indeed, in the case of rough symmetric objects, different cross-sections of the object observed from different poses of the US probe can have the same geometric properties. In this case, the minimization of the image features error does not guarantee the global convergence of the algorithm in terms of pose.

B. A new set of features using a multi-plane approach

Because of the geometry of the US sensor that provides information only in one plane, the major challenge of the US visual servoing is the control of the out-of-plane motions of the probe. On the contrary, the in-plane motions of the probe can be efficiently controlled using simple geometric features such as the coordinates of

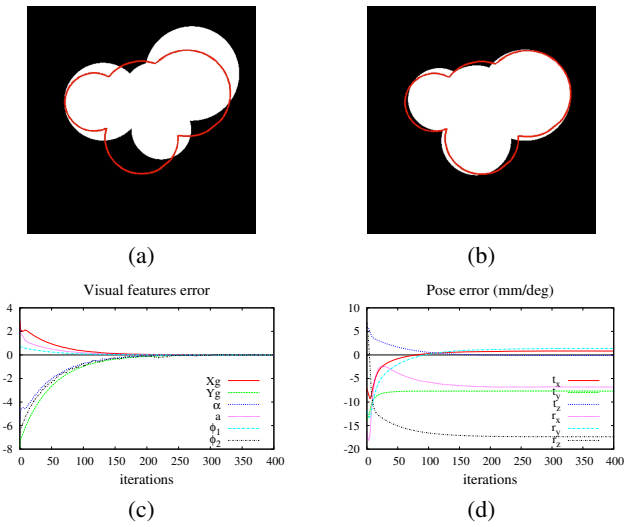

Fig. 1. Positioning task with mono-plane approach. (a,b) Initial and final cross-sections of the object (in white) with the desired contour superimposed (in red). The convergence in terms of visual error (c) does not correspond to the convergence in pose of the algorithm.

an image point for the translations and the main orientation of the section for the rotation. The general idea of our approach consists in selecting some geometric features strongly coupled to one particular motion of the probe in order to have an interaction matrix with good decoupling properties, which ensures an optimal trajectory of the probe. The visual features selected to control the in-plane motions of the probe are the coordinates (x_g, y_g) of the mass center of the object cross-section in the US image and its main orientation α:

$$\mathbf{s}_{in\text{-}plane} = (x_g, y_g, \alpha). \tag{5}$$

In order to control the six DOF of the US probe with these visual features, three orthogonal planes have to be considered. The Fig. 2 shows the tri-plane configuration we propose where the six visual features used to control both in-plane and out-of-plane motions of the probe are extracted from three orthogonal planes. Note that the 3 sections of the object of interest have to be fully visible in the 3 planes in order to compute their image moments. We define a control frame attached to the probe $\mathscr{F}_p$ and three frames $\mathscr{F}_{US_i}$ with $i \in \{0,1,2\}$ associated to the image planes. The plane US_0 is aligned with the plane $(\mathbf{x},\mathbf{y})$ of the probe, US_1 is aligned with the plane $(\mathbf{y},\mathbf{z})$ and US_2 is aligned with the plane $(\mathbf{x},\mathbf{z})$. In such a configuration, we can note that each motion of the probe corresponds to an in-plane motion in one of the three image planes (see Fig. 2). The in-plane velocities components (v_x, v_y, ω_z) of the probe correspond to the in-plane motions $(v_{x_0}, v_{y_0}, \omega_{z_0})$ of the plane US_0, its out-of-plane components (v_z, ω_x) correspond to the in-plane velocities $(v_{x_1}, -\omega_{z_1})$ of the plane US_1 and finally its out-of-plane rotation velocity ω_y corresponds to the in-plane rotation velocity $-\omega_{z_2}$ of the plane US_2. Therefore, we propose to control the probe with six image features coupled to in-plane motions of the image plane where they are defined. The image features vector that we retain is then:

$$\mathbf{s}_{multiplane} = (x_{g_0}, y_{g_0}, x_{g_1}, \alpha_1, \alpha_2, \alpha_0). \tag{6}$$

C. Interaction modeling

In each image plane US_i, the time variation of the moments-based image features $\mathbf{s}_i$ defined in (3) is related to the corresponding instantaneous velocity $\mathbf{v}_{c_i}$ according to:

$$\dot{\mathbf{s}}_i = \mathbf{L}_{\mathbf{s}_i}\, \mathbf{v}_{c_i} \, \forall i \in \{0,1,2\},$$

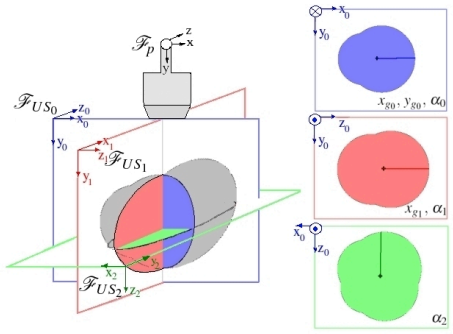

Fig. 2. The visual features are computed from three orthogonal planes. The probe frame coincides with the frame of US_0. On the right, this frame is reprojected in the various image plane frames. (Note that in this illustration the origins of the frames $\mathscr{F}_p$ and $\mathscr{F}_{US_i}$ are shifted for better visibility whereas they are superimposed in practice.)

where $\mathbf{L}_{\mathbf{s}_i}$ is the interaction matrix defined in (4).

In particular, each component of the features vector $\mathbf{s}_{multiplane}$ detailed in (6) is related to the velocity of its corresponding image plane as follows:

$$\begin{cases} \dot{x}_{g_0} &= \mathbf{L}_{x_{g_0}} \mathbf{v}_{c_0} \\ \dot{y}_{g_0} &= \mathbf{L}_{y_{g_0}} \mathbf{v}_{c_0} \\ \dot{x}_{g_1} &= \mathbf{L}_{x_{g_1}} \mathbf{v}_{c_1} \\ \dot{\alpha}_1 &= \mathbf{L}_{\alpha_1} \mathbf{v}_{c_1} \\ \dot{\alpha}_2 &= \mathbf{L}_{\alpha_2} \mathbf{v}_{c_2} \\ \dot{\alpha}_0 &= \mathbf{L}_{\alpha_0} \mathbf{v}_{c_0} \end{cases} \tag{7}$$

With the chosen configuration, the three planes frames are rigidly attached to the probe frame. We can therefore express the velocity $\mathbf{v}_{c_i}$ of each image plane in function of the instantaneous velocity of the probe $\mathbf{v}_c$:

$$\forall i \in \{0,1,2\}, \;\; \mathbf{v}_{c_i} = {}^i\mathbf{W}_p \, \mathbf{v}_c \tag{8}$$

with:

$${}^i\mathbf{W}_p = \begin{bmatrix} {}^i\mathbf{R}_p & [{}^i\mathbf{t}_p]_\times \, {}^i\mathbf{R}_p \\ 0_3 & {}^i\mathbf{R}_p \end{bmatrix} \tag{9}$$

Where ${}^i\mathbf{t}_p$ and ${}^i\mathbf{R}_p$ are the translation vector and the rotation matrix of the probe frame $\mathscr{F}_p$ expressed in the coordinate system of the image plane $\mathscr{F}_{US_i}$.

We obtain then after substituting (8) in (7) the interaction matrix that relates the variation of the features vector $\mathbf{s}_{multiplane}$ (6) to the motion of the probe (note that ${}^i\mathbf{t}_p = 0$ since frames $\mathscr{F}_p$ and $\mathscr{F}_{US_i}$ have same origin):

$$\mathbf{L}_{\mathbf{s}_{multiplane}} = \begin{bmatrix} -1 & 0 & x_{g_{0_{v_z}}} & x_{g_{0_{\omega_x}}} & x_{g_{0_{\omega_y}}} & y_{g_0} \\ 0 & -1 & y_{g_{0_{v_z}}} & y_{g_{0_{\omega_x}}} & y_{g_{0_{\omega_y}}} & -x_{g_0} \\ x_{g_{1_{v_z}}} & 0 & -1 & y_{g_1} & x_{g_{1_{\omega_y}}} & x_{g_{1_{\omega_x}}} \\ \alpha_{1_{v_z}} & 0 & 0 & 1 & \alpha_{1_{\omega_y}} & \alpha_{1_{\omega_x}} \\ 0 & \alpha_{2_{v_z}} & 0 & \alpha_{2_{\omega_x}} & 1 & \alpha_{2_{\omega_y}} \\ 0 & 0 & \alpha_{0_{v_z}} & \alpha_{0_{\omega_x}} & \alpha_{0_{\omega_y}} & -1 \end{bmatrix} \tag{10}$$

As stated previously, the six features chosen are coupled with one particular in-plane motion of their associated image plane. We propose then to relate their time variation only to the in-plane velocity components of their image frame. This means that we disregard the low variation of the image features due to the out-of-plane motions compared to the high variation due to in-plane motions. In the considered vector of image features $\mathbf{s}_{multiplane}$, the three parameters extracted from the first image are related to the in-plane velocity components of their image plane which coincide with the motions (v_x, v_y, ω_z) of the probe frame. In the image plane US_1, the x-coordinate of the mass center and the orientation of the object section

are used to control the in-plane velocity components (v_{x_1}, ω_{z_1}) that correspond to the components $(v_z, -\omega_x)$ in the US probe frame and the time variation of the object section orientation in the image plane US_2 is directly linked to the velocity component ω_y of the US probe. The approximated interaction matrix finally involved in the visual servoing control law (1) is then:

$$\widehat{\mathbf{L}}_{\mathbf{s}multiplane} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & y_{g_0} \\ 0 & -1 & 0 & 0 & 0 & -x_{g_0} \\ 0 & 0 & -1 & y_{g_1} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \tag{11}$$

This interaction matrix describes a simplified behavior of the system since the effect of the out-of-plane motions is neglected. Compared to the complete matrix given in (10), this one has great decoupling properties and is only dependent of the image features. In particular, the components of the estimated normal vector to the object surface are no longer involved.

D. Simulation validation and discussion

The geometrical simulator is now used to validate the multi-plane approach. The same initial and desired poses of the probe as in the previous mono-plane simulation presented in Fig. 1 are chosen and a virtual probe providing three orthogonal views is modeled in this simulator. The six geometric features (6) are computed from these US images and the control of the six DOF of the probe is performed using the estimated form of the interaction matrix (11). The results of one positioning task, where the control gain is $\lambda = 0.7$, are gathered in Fig. 3. The three internal views of the probe are displayed at its initial (a-c) and final (d-f) pose with the desired contour added in red. On the three final views of the probe the object cross section perfectly matches the desired contour, which validates visually the convergence of the task. Moreover we observe the convergence of each visual feature to its desired value on the curve (g). This visual convergence corresponds to the pose convergence of the probe as can be seen on the curve (h). The choice of six visual features extracted from three orthogonal images ensures a good behavior of the visual servoing with a convergence of the positioning task in terms of visual error and pose error. Moreover, with the approximated interaction matrix that neglects the effect of the out-of-plane motions on the chosen features, all the elements involved in the control law are directly measured in the current US image. It is expected that with a perfectly round shape, the algorithm will fall into a local minimum due to the ambiguity on the orientation features. However, in practice there is a low probability to encounter such round shape and small irregularities on the organ's shape should prevent such behavior.

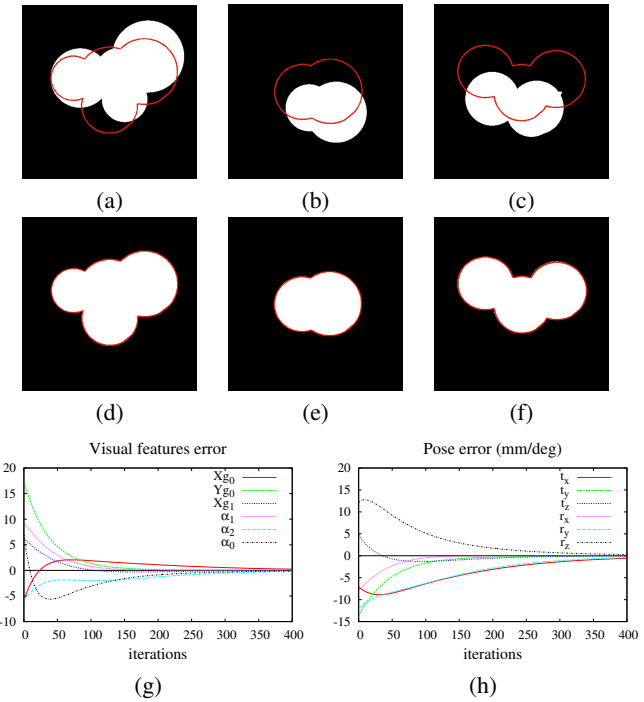

Fig. 3. Positioning task using six geometric features measured in three orthogonal images. (a-c) Initial object cross-section with the desired contour to reach (in red). (d-f) Observed object cross-sections at convergence of the algorithm. (g,h) Minimization of the visual and pose errors.

III. Real-time segmentation with Graph Cut algorithm

A graph cut segmentation [16] is chosen for its computational efficiency and its ability to generate binary segmentations with arbitrary topological properties. Pixels are represented as nodes of a graph. Every two neighboring pixels are connected by an edge (n-link) which cost is defined by pixel similarity. Additionally, each node is connected to two virtual terminal nodes. Cost of these edges (t-link) reflects the probability that the pixel belongs to foreground or background. A minimal cut of the graph then defines a labeling of each pixel as foreground or background. Segmentation of a sequence of US images follows these steps (details on implementation are in Section III-C):

1) Initialize boundaries of the tracked object (Section III-C);
2) Estimate the regional-probability model (Section III-B);
3) Proceed with the next image in the sequence;
4) Evaluate the n-links (Section III-A) and t-links (Section III-B) costs;
5) Perform the graph-cut segmentation and go back to 2.

A. Boundary constraints

The most frequently used method for calculating the boundary costs is the image gradient. This, however, does not work well with US images because of the noise and speckles. Instead, boundaries are identified between pixels with high phase congruency. This approach was introduced for US images by Mulet-Parada and Noble [17]. Phase congruency of 1D signals was detected in multiple directions in a 2D image with the use of log-Gabor filters. We use an undirected and computationally more efficient extension of this method [18]. Instead of using the oriented log-Gabor filters, the image is first filtered by Riesz's filters H_1 and H_2 which have the following representation in the Fourier domain:

$$H_1(u,v) = i\frac{u}{\sqrt{u^2+v^2}} \quad \text{and} \quad H_2(u,v) = i\frac{v}{\sqrt{u^2+v^2}}. \tag{12}$$

Then a log-Gabor filter g is applied yielding a monogenic signal f_M:

$$f_M(x,y) = \begin{bmatrix} f(x,y) * g(x,y), \\ f(x,y) * g(x,y) * h_1(x,y), \\ f(x,y) * g(x,y) * h_2(x,y) \end{bmatrix}, \tag{13}$$

where h_1 and h_2 are the Riesz's filters in the image domain (Eq. (12)), and g is defined in the Fourier domain as:

$$G(u,v) = \exp\left(-\frac{\left(\log(\sqrt{u^2+v^2}/\omega_0)\right)^2}{2\left(\log(k/\omega_0)\right)^2} \right). \tag{14}$$

The odd and even components of the monogenic signal f_M (Eq. (13)) are:

$$\begin{aligned} even_M(x,y) &= f_{M,1}(x,y), \\ odd_M(x,y) &= \sqrt{f_{M,2}^2 + f_{M,3}^2}, \end{aligned} \tag{15}$$

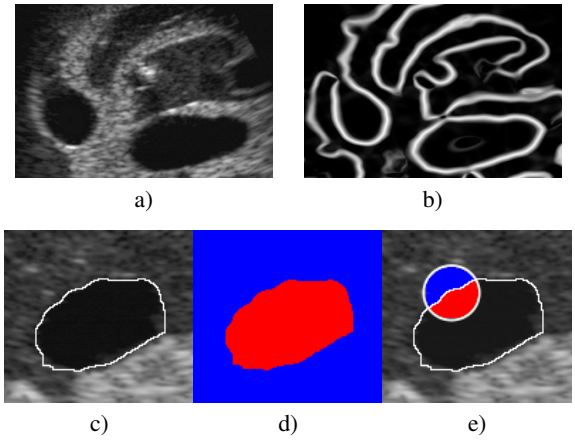

Fig. 4. (a) A cropped US image. (b) The corresponding phase-congruency measure (Eq. (16)). (c) The segmented image used for intensity probability model estimation. (d) The whole foreground and background regions are used to calculate the parameters of the global model (blue – foreground, red – background). (e) A small neighborhood of each of the boundary pixels is used for the local model estimation.

They are used to calculate the phase-congruency measure FA:

$$\begin{aligned} FA(x,y) &= \frac{\lfloor |odd_M(x,y)| - |even_M(x,y)| - T \rfloor}{\sqrt{odd_M^2(x,y) + even_M^2(x,y) + \varepsilon}}, \\ T &= \exp\left(\sum_{x,y} \frac{\log \sqrt{odd_M^2(x,y) + even_M^2(x,y)}}{N_x N_y} \right). \end{aligned} \tag{16}$$

where ε avoids division by zero, k controls the bandwidth and ω_0 is the filter center frequency. The following settings are used: $\omega_0 = 15$ pixels, $k/\omega_0 = 0.4$ and $\varepsilon = 10^{-5}$. Example of the phase-congruency measure is shown on Fig. 4(a-b). The n-link cost is then set to $exp(-FA(x,y))$.

B. Regional constraints

A global and a local probability model is used to calculate the costs of the t-links. Let us now assume that boundaries of an object are already segmented. For the global model, a mixture of Gaussians is used to model the pixel-intensity distribution in the foreground (2 Gaussians), and in the background (4 Gaussians), using intensities of pixels inside the object, and outside the object within a distance of 15 pixels, respectively. An Expectation-Maximization (EM) algorithm is used to estimate the parameters of the mixtures [19]. The parameters estimated for an object in one image are used to evaluate the edge costs in the successive image in the sequence. In US images, the foreground and especially the background are often not homogeneous. Therefore the global model is not optimal in all situations. For this reason, we have adopted a local model derived from the method proposed by Lankton and Tannenbaum [20]. A local intensity model is computed for each pixel on the object boundary using a circular neighborhood with a 15 pixels radius. The foreground and background parts of this neighborhood are used to estimate parameters of the local intensity model, see Fig. 4(c-e). A single Gaussian model is used instead of a mixture of Gaussian as applying the EM-algorithm for each boundary pixel would be too time consuming. For the t-link costs evaluation, the closest boundary pixel from the previous image is located for every pixel. Its Gaussian model is used to evaluate the foreground and background local probabilities.

C. Implementation

The segmentation is fully automatic with a semi-automatic initialization. The initialization is done by computing the phase-congruency measure (Eq. (16)) and splitting the image into several connected components. The object for tracking is then manually selected. The boundary and regional constraints and the graph cut segmentation are computed only in the close vicinity of the tracked object. The method is implemented in the CUDA language (Nvidia corporation, Santa Clara, California) allowing real-time computation on a dedicated graphic card. Parallelization of the t-links costs calculation is straightforward as it can be done for each pixel independently. Due to large filter size in the spatial domain, the convolution with Gabor and Riesz filters is done in the Fourier domain. The complete processing of one image plane (including segmentation and reinitialization of the intensity models) takes around 13.5 = ms for a single object (independent of the complexity of the object) fitting in a rectangular frame of up to 100x100 pixels. Loading the data and preprocessing takes on average 3.5 ms. The boundary and regional constraints are calculated in 2 ms and 0.4 ms, respectively. The graph-cut segmentation takes 6.1 ms and 1.5 ms is necessary for reinitialization of the object parameters.

IV. Robotic Experiments

Experiments have been performed on an ultrasound phantom, using a 6-DOF anthropomorphic robotic arm equipped with a motorized 3D probe and a force sensor (see Fig. 5). The 3D motorized ultrasound

Fig. 5. The anthropomorphic robotic system equipped with a motorized 3D US probe and the ultrasound phantom.

probe (model: 4DC7-3, Ultrasonix Company) is attached to the end-effector of the robot (model Viper 850, Adept Company) and is connected to an ultrasound imaging workstation (SonixTouch, Ultrasonix Company) that grabs 3D volumes at a rate close to 2 volumes/second. Each volume is composed of 27 slices and exhibits a field of view around the scanning motor axis of 40 deg with a depth of 12 cm. We use an ultrasound phantom (model 55 from CIRS company) that is usually employed for ultrasound calibration purpose to simulate soft-tissues and an organ of interest. It contains an egg-shaped object with a clear axial symmetry. A force sensor is fixed between the robot end-effector and the probe to measure force interaction between the probe and the phantom. The image processing and the control law computation are performed on a PC equiped with a Dual-core 2.4 Ghz Intel Pentium and GPU. The same force controller as the one implemented in [12] is applied to control the translation velocity along the y-axis of the probe frame in such a way to regulate the contact force to 1 N. The remaining 5 DOF of the system are controlled by visual servoing.

The experiment consists in automatically positioning the 3D ultrasound probe with respect to the egg-shaped object contained in the phantom in such a way to reach desired sections observed in the three orthogonal image planes. In this experiment, we test our new multi-plane visual servoing approach described in section II-B. The six visual features $\mathbf{s}_{multiplane}$ are extracted thanks to the graph-cut segmentation algorithm presented in section III.

We first tele-operate the robot to position the probe to a desired location where the desired visual features are saved. Then we move the probe away to another location that we consider as being the initial pose. The measured initial pose error before launching the

visual servoing is: $\Delta\mathbf{r}_{init}(mm,deg) = [-18,1,17,-8.8,-20.5,-7.8]$. We apply our multi-plane visual servoing approach based on the approximated interaction matrix (11). During the visual servoing process the control gain was set to $\lambda = 0.2$ and the control velocity sent to the probe was updated every 80 ms (even though a single-volume acquisition time is around 500 ms). The supplementary video material accompanying this paper shows the experiment. Fig. 6(a)-(c) presents the observed ultrasound images in the 3 orthogonal planes before launching the visual servoing. The green contours correspond to the initial and current sections and the red ones display the contours of the desired sections. The three ultrasound images obtained at convergence are reported in Fig. 6(d)-(f) and demonstrate that the desired sections are correctly reached. The time evolution of the visual error and the pose error are reported on Fig. 7. One can observe that the convergence to zero is obtained both for the visual and pose error after 20 seconds. The final pose error measure gives: $\Delta\mathbf{r}_{final}(mm,deg) = [-0.5,0.1,0.3,-0.24,0.05,-0.39]$. These results experimentally demonstrate that this multi-plane approach is appropriate for positioning application with respect to objects exhibiting strong symmetry in opposite to the mono-plane approach that fails in this case. Object motion compensation is however limited by the low volume rate of our motorized 3D US probe. Nevertheless, as the segmentation process takes only $3 \times 13.5 = 40.5$ ms for the three US planes, automatic compensation could be performed with the use of a matrix array 3D US probe that provides 25 volumes/s.

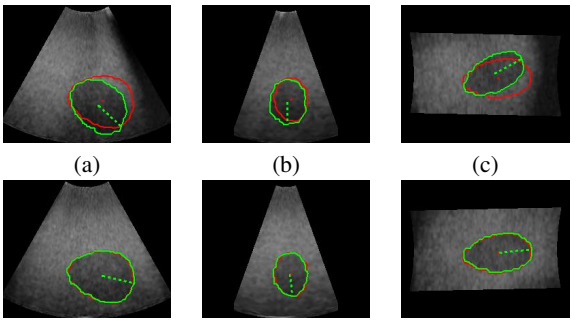

Fig. 6. Positioning task with six features extracted from three orthogonal images. (a-c) Initial object cross-section (in green) with the desired contour to reach (in red). (d-f) Observed object cross-sections at convergence of the visual servoing.

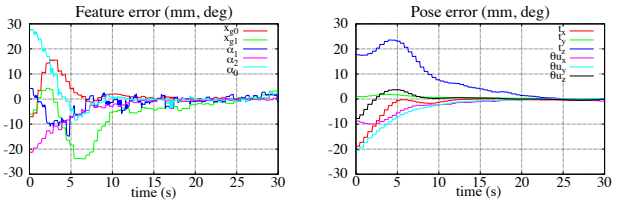

Fig. 7. Positioning task with six features extracted from three orthogonal images. Decrease of the visual error (left) and pose error (right) during the visual servoing.

V. Conclusion

In this paper, a new approach is proposed to control the full motions of a robotized ultrasound probe by image-based visual servoing using six geometric features computed from three orthogonal planes. The considered visual features vector characterizes efficiently the pose of the probe with respect to the target and the choice of features strongly coupled to in-plane motions of the image allows to neglect the effect of out-of-plane motions and therefore to model an approximated interaction matrix whose all elements can be measured in the US images. Moreover in order to extract the visual features a fast image processing algorithm based on a graph cut strategy is proposed to segment in real-time the contour of the object of interest observed in the three orthogonal planes. With the phase-congruency measure, robust segmentation of US is achieved even in the presence of noise and speckles. The local intensity probability model will allow the algorithm to be used also in more complex phantoms and in real patients.

References

[1] F. Chaumette and S. Hutchinson, Visual Servo Control, Part I: Basic Approaches. *IEEE Robotics and Automation Magazine* vol. 13(4), pp. 82-90, 2006.

[2] 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), pp. 11-23, 2002.

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

[4] 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*, pp. 122-129, 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. *IEEE Int. Conf. on Robotics and Automation*, pp. 2655-2660, 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. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp. 877-882, 2007.

[7] J.A. Stoll, P.M. Novotny, R.D. Howe and P.E. Dupont, Real-time 3D Ultrasound-based Servoing of a Surgical Instrument. *In IEEE Int. Conf. on Robotics and Automation*, pp. 613-618, Orlando, USA, 2006.

[8] 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*, pp. 3021-3027, Roma, Italy, 2007.

[9] M. Sauvee, P. Poignet, E. Dombre, Ultrasound image based visual servoing of a surgical instrument through non-linear model predictive control. *Int. Journal of Robotics Research*, vol. 27(1), pp. 25-40, 2008.

[10] 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*, pp. 5267 - 5272, Shanghai, China, 2011.

[11] 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), pp. 1334-1354, 2009.

[12] C. Nadeau and A. Krupa. Intensity-based ultrasound visual servoing: modeling and validation with 2D and 3D probes. *IEEE. Trans. on Robotics*, vol. 29(4), pp. 1003-1015, 2013.

[13] 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), pp. 296-306, 2010.

[14] 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*, pp. 5706 - 5711, Taipei, Taiwan, 2010.

[15] F. Chaumette, Image moments: A general and useful set of features for visual servoing. *IEEE Trans. on Robotics*, vol. 20(4), pp. 713-723, 2004.

[16] Y. Boykov and V. Kolmogorov, An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. On Pattern Analysis And Machine Intelligence*, vol. 26(9), pp. 1124-1137, 2004.

[17] M. Mulet-Parada and J.A. Noble, 2D+T acoustic boundary detection in echocardiography. *Medical image analysis*, vol. 4, pp. 21-30, 2000.

[18] K. Rajpoot, V. Grau and J.A. Noble, Local-phase based 3D boundary detection using monogenic signal and its application to real-time 3-D echocardiography images. *IEEE int. conf. on Symposium on Biomedical Imaging: From Nano to Macro*, pp. 783 - 786, 2009.

[19] A.P. Dempster, N.M. Laird and D.B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39(1), pp. 1-38, 1977.

[20] S. Lankton and A. Tannenbaum, Localizing region-based active contours. *IEEE Trans. On Image Processing*, vol. 17(11), pp. 2029-2039, 2008.