

# Joint Co-segmentation and Registration of 3D Ultrasound Images

Raphael Prevost<sup>1,2</sup>, Remi Cuingnet<sup>1</sup>, Benoit Mory<sup>1</sup>, Jean-Michel Correas<sup>3</sup>,  
Laurent D. Cohen<sup>2</sup>, and Roberto Ardon<sup>1</sup>

<sup>1</sup> Philips Research Medisys, Suresnes, France

<sup>2</sup> CEREMADE UMR 7534, Université Paris Dauphine, Paris, France

<sup>3</sup> Adult Radiology Department, Necker Hospital, Paris, France

**Abstract.** Contrast-enhanced ultrasound (CEUS) allows a visualization of the vascularization and complements the anatomical information provided by conventional ultrasound (US). However, these images are inherently subject to noise and shadows, which hinders standard segmentation algorithms. In this paper, we propose to use simultaneously the different information coming from 3D US and CEUS images to address the problem of kidney segmentation. To that end, we introduce a generic framework for joint co-segmentation and registration that seeks objects having the same shape in several images. From this framework, we derive both an ellipsoid co-detection and a model-based co-segmentation algorithm. These methods rely on voxel-classification maps that we estimate using random forests in a structured way. This yields a fast and fully automated pipeline, in which an ellipsoid is first estimated to locate the kidney in both US and CEUS volumes and then deformed to segment it accurately. The proposed method outperforms state-of-the-art results (by dividing the kidney volume error by two) on a clinically representative database of 64 images.

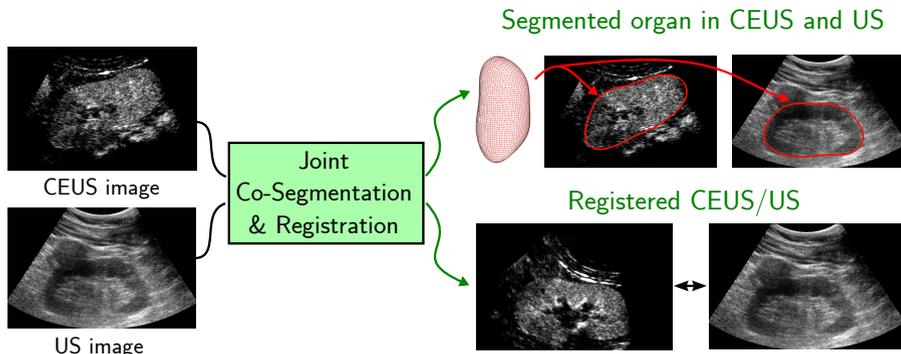
**Keywords:** co-segmentation, registration, kidney, random forests, ultrasound, contrast-enhanced ultrasound.

## 1 Introduction

### 1.1 Clinical Setting

Contrast-enhanced ultrasound (CEUS) consists in acquiring a ultrasound image after injecting in the patient's blood a contrast agent made of gas-filled microbubbles. Since those bubbles have a different acoustic response from the tissues, they can be isolated and images showing only the blood flow can be generated [1]. This modality is particularly valuable for visual assessment of the functioning of highly vascularized organs like kidneys. Yet, analysis of such images can be very challenging and literature on their segmentation is limited.

In [13], we proposed a method to detect and segment kidneys in 3D CEUS images. While we provided an automated pipeline, failures were reported in several cases and user interactions were needed to obtain a satisfying result. Yet,



**Fig. 1.** Joint co-segmentation and registration. Given two different non-aligned images of the same object, the proposed method aims at segmenting this object in both images as well as estimating a rigid transformation between them.

because of shadowing effects, pathologies and restricted field of view, parts of the kidney may be hardly visible in the image. In such cases even expert users may have difficulty delineating the true boundary of the organ by solely relying on the CEUS images. In clinical routine every CEUS acquisition is preceded by a conventional US acquisition to locate the kidney. Hence, the latter could be used to complement the CEUS image and thus cope with missing and corrupted information. However, automated kidney segmentation in 3D US images is also an open issue. Martin-Fernandez and Alberola-Lopez [8] tackled this problem but their method requires a manual initialization. For both US and CEUS segmentation are equally challenging, we propose to address them simultaneously by performing kidney co-segmentation in the two images.

## 1.2 Related Work on Co-segmentation and Registration

Co-segmentation often denotes the task of finding an object in each image that shares the same appearance but not necessarily the same shape [16]. Here we look for the exactly same organ in two images but with a different appearance. As simultaneous acquisition of US and CEUS is not possible on current 3D imaging systems, the two images are in arbitrary referentials and need to be aligned. However, standard iconic registration methods are not adapted since visible structures, apart from the kidney itself, are completely different in US and CEUS. Co-segmentation shall therefore help registration, just as registration helps co-segmentation. This calls for a method that jointly performs these two tasks (see Figure 1).

Although segmentation and registration are often seen as two separate problems, several approaches have already been proposed to perform them

simultaneously. Most of them rely on an iconic registration guiding the segmentation (e.g. [17,12,7]). Yet they assume that the segmentation is known in one of the images, which is not the case in our application of co-segmentation. Moreover, as stated before, CEUS/US intensity-based registration is bound to fail since visible structures does not correspond to each other. Instead of registering the images themselves, Wyatt et al. [18] developed a MAP formulation to perform registration on label maps resulting from a segmentation step. However no shape model is enforced and noise can degrade the results. In [19], Yezzi et al. introduced a variational framework that consists in a feature-based registration in which the features are actually the segmenting active contours.

In this paper, we aim at extending both the kidney detection and segmentation in a 3D CEUS image presented in [13] to a pair of 3D CEUS and US images. To that end, we develop a generic joint co-segmentation and registration framework inspired by [19]. This results in a fully automated pipeline to obtain both an improved kidney segmentation in CEUS and US images and a registration of them.

The article is structured as follows. Section 2 describes the generic framework and its application to two consecutive algorithms. Both rely on an appearance characterization of the kidney in ultrasound images that is learnt using random forest in an original structured way (Section 3). Results of the proposed co-segmentation method on a challenging clinical database are presented in Section 4. Finally, Section 5 provides some discussion and concludes the paper.

## 2 Joint Co-segmentation and Registration

### 2.1 Generic Implicit Variational Framework

Segmentation consists in finding an optimal two-phase (inside and outside) partitioning of a given image  $I : \Omega \rightarrow \mathbb{R}^+$ . In implicit methods, this partitioning is defined using the sign of an implicit function  $\phi : \Omega \rightarrow \mathbb{R}$ . In [13], two variational methods are developed to respectively detect and segment the kidney. They both consist in seeking  $\phi$  as the minimum of functional of the following generic form

$$E_I(\phi) = \int_{\Omega} f(\phi(\mathbf{x})) r_I(\mathbf{x}) d\mathbf{x} + \mathcal{R}(\phi) \quad (1)$$

where  $f$  is a real-valued function and  $r_I(\mathbf{x})$  denotes a pointwise score on whether  $\mathbf{x}$  looks like an interior or exterior voxel in the image  $I$ . This is a standard setting in which the optimal implicit function  $\phi$  must achieve a trade-off between an image-based term and a regularization term  $\mathcal{R}$ . For example, the seminal method of Chan and Vese [3] falls in this framework with  $f = H$  the Heaviside function and  $r_I(\mathbf{x}) = (I(\mathbf{x}) - c_{int})^2 - (I(\mathbf{x}) - c_{ext})^2$  with  $c_{int}$  and  $c_{ext}$  denoting mean intensities inside and outside the target object.

We are interested in the case where a pair of images  $I_1 : \Omega_1 \rightarrow \mathbb{R}$  and  $I_2 : \Omega_2 \rightarrow \mathbb{R}$  of the same object are available. If those images were perfectly aligned, the energy in Eq (1) can be straightforwardly generalized to perform co-segmentation:

$$E_{I_1, I_2}(\phi) = \int_{\Omega_1} f(\phi(\mathbf{x})) (r_{I_1}(\mathbf{x}) + r_{I_2}(\mathbf{x})) d\mathbf{x} + \mathcal{R}(\phi). \quad (2)$$

Unfortunately, such an assumption rarely holds in medical applications unless the two images are acquired simultaneously. A more realistic hypothesis is to assume that the target object, segmented by  $\phi$ , is not deformed between the two acquisitions, but only undergoes an unknown rigid transformation  $\mathcal{G}_r$ . The co-segmentation energy thus reads

$$E_{I_1, I_2}(\phi, \mathcal{G}_r) = \int_{\Omega_1} f(\phi(\mathbf{x})) r_{I_1}(\mathbf{x}) d\mathbf{x} + \int_{\Omega_2} f(\phi \circ \mathcal{G}_r(\mathbf{x})) r_{I_2}(\mathbf{x}) d\mathbf{x} + \mathcal{R}(\phi). \quad (3)$$

Note that, after a variable substitution, it can be equivalently written

$$E_{I_1, I_2}(\phi, \mathcal{G}_r) = \int_{\Omega_1} f(\phi(\mathbf{x})) (r_{I_1}(\mathbf{x}) + r_{I_2} \circ \mathcal{G}_r^{-1}(\mathbf{x})) d\mathbf{x} + \mathcal{R}(\phi). \quad (4)$$

Minimizing  $E_{I_1, I_2}$  with respect to  $\phi$  and  $\mathcal{G}_r$  simultaneously can be therefore interpreted as performing jointly segmentation (via  $\phi$ ) and rigid registration (via  $\mathcal{G}_r$ ). This generalizes a more common co-segmentation approach (e.g. [5]) where the images are first aligned in a preprocessing step. Note that for clarity, we only consider two images but all equations can be generalized straightforwardly to an arbitrary number of images.

In the following, we apply this framework to (i) a robust ellipsoid detection [13] and (ii) implicit template deformation [10] to build a completely automated workflow for kidney segmentation in CEUS and US images. Note that the kidney, which is surrounded by a tough fibrous renal capsule, is a rigid organ. The hypothesis of non-deformation is therefore justified.

## 2.2 Robust Ellipsoid Co-detection

In [13], we proposed to detect the kidney in CEUS images as an ellipsoid. For that purpose, we developed a variational framework to achieve fast and robust ellipsoid detection. Any ellipsoid can be implicitly represented by a function  $\phi_{\mathbf{c}, \mathbf{M}} : \Omega \rightarrow \mathbb{R}$  such that  $\phi_{\mathbf{c}, \mathbf{M}}(\mathbf{x}) = 1 - (\mathbf{x} - \mathbf{c})^T \mathbf{M} (\mathbf{x} - \mathbf{c})$ , where  $\mathbf{c} \in \mathbb{R}^3$  denotes the ellipsoid center and  $\mathbf{M}$  is a symmetric positive-definite matrix. The ellipsoid interior is then the zero superlevel set of  $\phi_{\mathbf{c}, \mathbf{M}}$ . Given a probability map  $p : \Omega \rightarrow [0, 1]$  of the target object, defined at each pixel, the detection is sought as the smallest ellipsoid that includes most of the pixels  $\mathbf{x}$  with high probability  $p(\mathbf{x})$ . To limit the influence of possible false positives pixels, a weighting function  $w : \Omega \rightarrow [0, 1]$  acting on  $p$  is also estimated. We thus proposed to solve the following problem

$$\begin{aligned} \min_{\mathbf{c}, \mathbf{M}, w} E_{det}(\mathbf{c}, \mathbf{M}, w) &= - \int_{\Omega} \phi_{\mathbf{c}, \mathbf{M}}(\mathbf{x}) p(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} \\ &+ \mu \cdot \left( \int_{\Omega} p(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} \right) \cdot \log \left( \frac{\mathcal{V}ol(\mathbf{M})}{|\Omega|} \right) \end{aligned} \tag{5}$$

with  $\mathcal{V}ol(\mathbf{M}) = \frac{4\pi}{3} \sqrt{\det \mathbf{M}^{-1}}$  the ellipsoid volume.

Such a setting falls into the framework described in Eq (1) :

- with  $f = Id$  and  $r_I = -pw$  in the image-based term.  $r_I$  is then highly negative at voxels that have a high probability and are not outliers. To minimize the energy, such pixels must be inside the ellipsoid i.e. where  $\phi$  is positive.
- with  $\mathcal{R}(\phi_{\mathbf{c}, \mathbf{M}}) = \mathcal{R}(\mathbf{M}) = \mu \cdot \int_{\Omega} pw \cdot \log \left( \frac{\mathcal{V}ol(\mathbf{M})}{|\Omega|} \right)$  as a regularization term that penalizes the volume of the ellipsoid. The rationale behind the logarithm is statistical: the energy in Eq (5) is closely related to maximum likelihood estimation of a Gaussian distribution. Factor  $\int_{\Omega} pw$  normalizes the contribution of such a term, while  $\mu$  denotes a trade-off parameter set to  $\frac{1}{2}$  in 2D and  $\frac{2}{3}$  in 3D (see [13]).

Expanding this algorithm to another image with a given probability  $p_2$  requires the introduction of another weighting function  $w_2$ . Following Eq (3), we can now define the co-detection energy as

$$\begin{aligned} E_{co-det}(\mathbf{c}, \mathbf{M}, w_i, \mathcal{G}_r) &= - \int_{\Omega} \phi_{\mathbf{c}, \mathbf{M}} p_1 w_1 - \int_{\Omega} \phi_{\mathbf{c}, \mathbf{M}} \circ \mathcal{G}_r p_2 w_2 \\ &+ \mu \left( \int_{\Omega} p_1 w_1 + p_2 w_2 \right) \log \left( \frac{\mathcal{V}ol(\mathbf{M})}{|\Omega|} \right) \end{aligned}$$

with  $\mathcal{V}ol(\mathbf{M}) = \frac{4\pi}{3} \sqrt{\det \mathbf{M}^{-1}}$  the ellipsoid volume. (6)

To facilitate the resolution of such a problem,  $\mathcal{G}_r$  - as a rigid transformation - can be decomposed into a rotation and a translation. We can therefore equivalently write the energy as a function of the ellipsoid center  $\mathbf{c}_2$  in the second image and the rotation matrix  $\mathbf{R}$  :

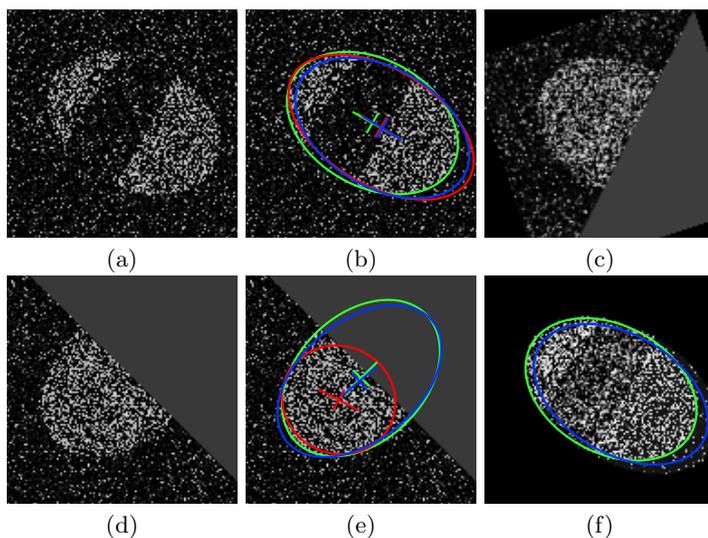
$$\begin{aligned} E_{co-det}(\mathbf{c}_i, w_i, \mathbf{R}, \mathbf{M}) &= - \int_{\Omega} \phi_{\mathbf{c}_1, \mathbf{M}}(\mathbf{x}) p_1(\mathbf{x}) w_1(\mathbf{x}) d\mathbf{x} \\ &- \int_{\Omega} \phi_{\mathbf{c}_2, \mathbf{R}^T \mathbf{M} \mathbf{R}}(\mathbf{x}) p_2(\mathbf{x}) w_2(\mathbf{x}) d\mathbf{x} \\ &+ \mu \left( \int_{\Omega} p_1 w_1 + p_2 w_2 \right) \log \left( \frac{\mathcal{V}ol(\mathbf{M})}{|\Omega|} \right) \end{aligned} \tag{7}$$

Minimization of such functional is done in an alternate three-step process:

1. The statistical interpretation still holds for the ellipsoids centers and matrix: minimizers  $\mathbf{c}_1^*$  and  $\mathbf{c}_2^*$  are weighted centroids while minimizer  $\mathbf{M}^*$  is related to the weighted covariance matrix of pixels coming from both images.

2. The unknown matrix  $\mathbf{R}$  accounts for the possible rotation between the two images and can be parametrized by a vector of angles  $\Theta \in \mathbb{R}^3$ . A gradient descent is performed at each iteration to minimize the energy with respect to  $\Theta$ .
3. The weights  $w_1$  and  $w_2$  are finally updated as indicator functions (up to a slight dilation) of the current ellipsoid estimates.

The complete minimization strategy is summarized in Algorithm 1. This algorithm is computationally efficient : closed-form solutions are available (except for  $\mathbf{R}$ ) and the process, though iterative, usually converges in very few iterations.



**Fig. 2.** Ellipse detection on two synthetic images with  $p_1$  (a) and  $p_2$  (d). Detected ellipses with their center and main axes are shown in (b) and (e) for independent ellipse detection (red) and proposed method for co-detection (blue) compared to the ground truth (green). (c) Second image registered with the estimated transform  $\mathcal{G}_r^{-1}$ . (f) Combination of image terms  $w_1 p_1 + (w_2 p_2) \circ \mathcal{G}_r^{-1}$  used for ellipse estimation at convergence.

Figure 2 shows an example of ellipse co-detection in synthetic images, where the probability of belonging to the target object is the image intensity. Despite the noise, the simulated shadow and the reduced field-of-view effect, the co-detection algorithm provides a good estimate on the ellipse position, size and orientation in both images.

### 2.3 Co-segmentation via Implicit Template Deformation

The previously detected ellipsoid is not a precise segmentation of the kidney, but can be used as an initialization for a more elaborate segmentation method, namely template deformation [14,10].

**Algorithm 1.** Robust ellipsoid co-detection algorithm

---

```

initialization  $\forall \mathbf{x} \in \Omega, w_1(\mathbf{x}) \leftarrow 1, w_2(\mathbf{x}) \leftarrow 1$ 
repeat
  // Estimation of centers  $\mathbf{c}_1$  and  $\mathbf{c}_2$  and matrix  $\mathbf{M}$ 
   $\mathbf{c}_1 \leftarrow \frac{1}{\int_{\Omega} p_1 w_1} \int_{\Omega} p_1(\mathbf{x}) w_1(\mathbf{x}) \mathbf{x} d\mathbf{x}$ 
   $\mathbf{c}_2 \leftarrow \frac{1}{\int_{\Omega} p_2 w_2} \int_{\Omega} p_2(\mathbf{x}) w_2(\mathbf{x}) \mathbf{x} d\mathbf{x}$ 
   $\mathbf{M}^{-1} \leftarrow \frac{2}{\mu \int_{\Omega} p_1 w_1 + p_2 w_2} \left( \int_{\Omega} p_1(\mathbf{x}) w_1(\mathbf{x}) (\mathbf{x} - \mathbf{c}_1) (\mathbf{x} - \mathbf{c}_1)^T d\mathbf{x} \right.$ 
     $\left. + \int_{\Omega} p_2(\mathbf{x}) w_2(\mathbf{x}) \mathbf{R} (\mathbf{x} - \mathbf{c}_2) (\mathbf{x} - \mathbf{c}_2)^T \mathbf{R}^T d\mathbf{x} \right)$ 
  // Update of the rotation matrix  $\mathbf{R}$  by gradient descent with time step  $\Delta t$ 
  repeat
    |  $\mathbf{R}(\Theta) \leftarrow \mathbf{R}(\Theta - \Delta t \nabla_{\Theta} E_{co-det}(\Theta))$ 
  until convergence;
  // Update of the weighting functions  $w_1$  and  $w_2$  for each  $\mathbf{x} \in \Omega$ 
  if  $(\mathbf{x} - \mathbf{c})^T \mathcal{M}(\mathbf{x} - \mathbf{c}) \leq 1 - \mu \log\left(\frac{\text{Vol}(\mathbf{M})}{|\Omega|}\right)$  then
    |  $w_1(\mathbf{x}) \leftarrow 1$  else  $w_1(\mathbf{x}) \leftarrow 0$ 
  if  $(\mathbf{x} - \mathbf{c}_2)^T \mathbf{R}^T \mathbf{M} \mathbf{R} (\mathbf{x} - \mathbf{c}_2) \leq 1 - \mu \log\left(\frac{\text{Vol}(\mathbf{M})}{|\Omega|}\right)$  then
    |  $w_2(\mathbf{x}) \leftarrow 1$  else  $w_2(\mathbf{x}) \leftarrow 0$ 
until convergence;

```

---

Template deformation is a model-based segmentation framework that represents the segmented object as a deformed initial function (called template). In an implicit setting [10], this segmentation is represented by the zero-level set of a function  $\phi : \Omega \rightarrow \mathbb{R}$  defined as  $\phi = \phi_0 \circ \psi$ , where  $\phi_0$  is the implicit template and the transformation  $\psi : \Omega \rightarrow \Omega$  becomes the unknown of the problem.  $\psi$  is sought as a minimum of the following energy

$$E_{seg}(\psi) = \int_{\Omega} H(\phi_0 \circ \psi) r_I(\mathbf{x}) d\mathbf{x} + \mathcal{R}(\psi) . \quad (8)$$

where  $H$  is the Heaviside function (i.e.  $H(x) = 1$  if  $x > 0$ , otherwise 0) and  $r_I$  an image-based term negative (resp. positive) at pixels likely to be inside (resp. outside) the target object. The template  $\phi_0$  acts as a shape prior and the transformation  $\psi$  that  $\phi_0$  undergoes is penalized via  $\mathcal{R}$ . In order to define this regularization term, this transformation is decomposed as  $\psi = \mathcal{L} \circ \mathcal{G}$  where

- $\mathcal{G}$  is a global transformation that accounts for the pose and scale of the segmentation. It is defined through a vector of parameters (typically in  $\mathbb{R}^7$  for a 3D similarity);
- $\mathcal{L}$  is a non-rigid local deformation, expressed using a displacement field  $\mathbf{u}$  such that  $\mathcal{L}(\mathbf{x}) = \mathbf{x} + (\mathbf{u} * K_{\sigma})(\mathbf{x})$ .  $K_{\sigma}$  is a Gaussian kernel that provides built-in smoothness.

This decomposition allows  $\mathcal{R}$  to be pose-invariant and constrains only the non-rigid deformation :  $\mathcal{R}(\psi) = \mathcal{R}(\mathcal{L}) = \int_{\Omega} \|\mathcal{L} - Id\|^2 = \int_{\Omega} \|\mathbf{u} * K_{\sigma}\|^2$ . Penalizing

the magnitude of the displacement field allows to control the deviation of the segmentation from the initial shape prior.

Implicit template deformation, as previously described, is part of the framework defined in Eq. (1) with  $f = H$ . We can therefore extend it to co-segmentation using Eq. (3) by considering the following functional:

$$E_{co-seg}(\mathcal{L}, \mathcal{G}, \mathcal{G}_r) = \int_{\Omega} H(\phi_0 \circ \mathcal{L} \circ \mathcal{G}) r_{I_1}(\mathbf{x}) d\mathbf{x} + \int_{\Omega} H(\phi_0 \circ \mathcal{L} \circ \mathcal{G} \circ \mathcal{G}_r) r_{I_2}(\mathbf{x}) d\mathbf{x} + \frac{\lambda}{2} \|\mathcal{L} - Id\|_2^2. \quad (9)$$

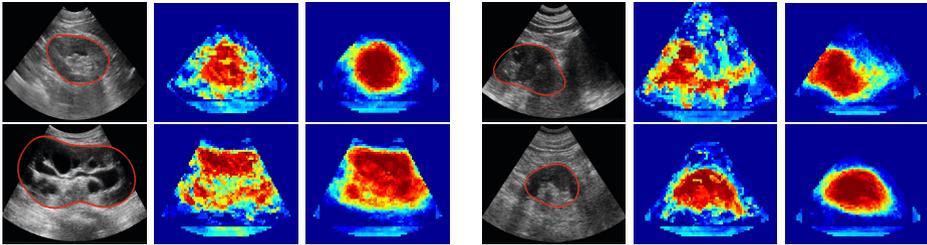
In our application, the template  $\phi_0$  is defined as the implicit representation of the detected ellipsoid  $\phi_{\mathbf{c}_1, \mathcal{M}}$ .  $\mathcal{G}$  and  $\mathcal{L}$  are initially set to the identity while  $\mathcal{G}_r$  is initialized with the previously estimated registering transformation:  $\mathcal{G}_r(\mathbf{x}) = \mathbf{R}(\mathbf{x} + \mathbf{c}_1 - \mathbf{c}_2)$ . As in [13], the image-terms are defined as  $r_{I_i} = \Delta p_i$  where  $\Delta$  denotes the Laplacian operator and  $p_i$  the kidney probability in image  $i$ . The energy  $E_{co-seg}$  is then minimized with respect to the parameters of  $\mathcal{G}$ ,  $\mathcal{G}_r$  and each component of the vector field  $\mathbf{u}$ , through a gradient descent.

### 3 Learning Kidney Appearance Using Random Forests

The previously described algorithms rely on functions  $p_I$  that associate to each voxel  $\mathbf{x}$  of the image  $I$  a probability to belong to the kidney. In CEUS images, bright areas indicate the presence of contrast agent which is mainly localized in the kidney. Therefore we can directly use the normalized intensity of the image as a probability term, i.e.  $p_{CEUS} = \frac{I_{CEUS}}{\max I_{CEUS}}$ .

However, the kidney appearance has a much higher variability in US images, although their structure is consistent: kidneys are always composed of a bright sinus surrounded by a darker parenchyma (see Figure 3). As intensity itself is not reliable enough, we chose to combine multiple image features using decision forests [2] to obtain a class posterior map  $p_{US}$ . Recent work [11,9,6,4,20] demonstrated that adding contextual information allows to improve spatial consistency and thus classification performance. Here we propose to exploit the kidney structure in a simple yet efficient way. Similarly to the auto-context framework introduced by Tu et al. [15], contextual information is included by using two classifiers in cascade. A first classification (kidney vs background) is performed in each voxel using a decision forest. Then we use these class posterior probabilities as additional input of a second random forest that will give the final kidney probability  $p_{US}$ .

The features used for the first decision forest were the intensity of the image and its Laplacian at the considered voxel as well as at its neighbors' within a  $7 \times 7 \times 7$  local patch, at three different scales ( $\sigma = 2, 4, 6$  mm). Intensities were normalized in each patch. For the second forest, we added the estimated class posterior as additional channels. Each forest was composed of 10 trees with maximum depth 15.



**Fig. 3.** Kidney appearance in US images (denoted in red). Left: original images showing the high variability of the database. Middle: kidney probability given by the first classifier. Right: final kidney probability  $p_{US}$ .

## 4 Experiments and Validation

### 4.1 Material

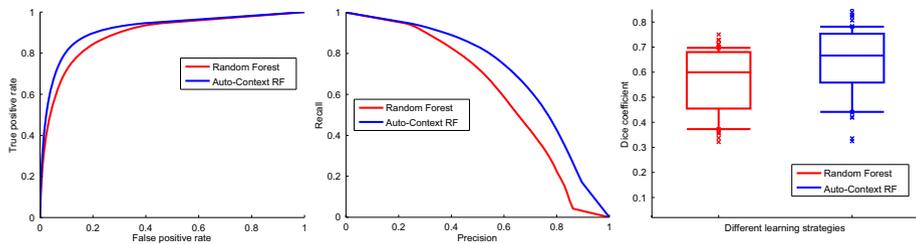
Our database is composed of 64 couples of CEUS and US volumes acquired from 35 different patients. This set is clinically representative as different ultrasound probes were used, with different fields of view, on both diseased and healthy kidneys. The volumes size was  $512 \times 510 \times 256$  voxels with varying spatial resolutions ( $0.25 \times 0.25 \times 0.55$  mm in average). The CEUS acquisitions have been performed a few seconds after injection of 2.4 mL of Sonovue (Bracco, Italy) contrast agent. Kidney segmentation made by an expert was available for each image as a ground truth. The proposed method was implemented in C++ and the average overall computational time was around 20 seconds on a standard computer (Intel Core i5 2.67 Ghz, 4GB RAM).

### 4.2 Validation on the Learnt Kidney Appearance in US Images

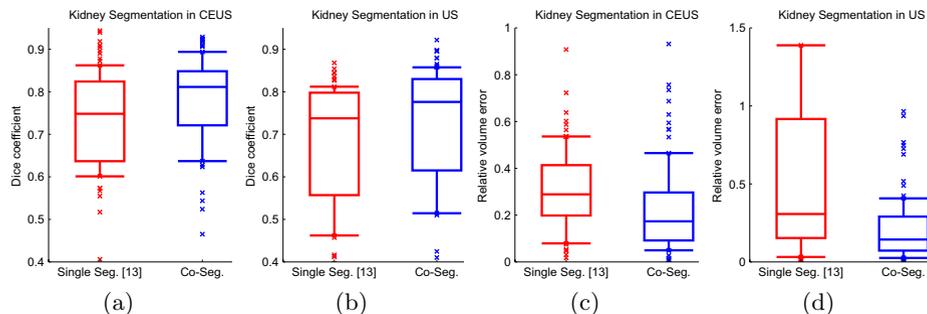
The patient database was split into two groups. Results on the whole dataset were then obtained using a two-fold cross-validation. Figure 4 shows the ROC and Precision-Recall curves, as well as a boxplot of the Dice coefficients obtained by thresholding the kidney probabilities computed (i) by the first decision forest and (ii) using the auto-context approach with another forest in cascade. The latter provides better kidney probabilities with respect to all reported statistics. In particular, Dice coefficients are significantly improved, with a p-value  $< 10^{-4}$  (in this paper, all p-values were obtained for the Wilcoxon signed-rank test after Bonferroni correction). Indeed, taking into account structural information helps for example in distinguishing the kidney sinus from the background or the parenchyma from shadows, and allows a more spatially coherent classification (see Figure 3).

### 4.3 Validation on the Kidney Co-segmentation

In all CEUS/US couples, kidneys were co-detected using Section 2.2 as an initialization for the co-segmentation algorithm of Section 2.3. For comparison, we

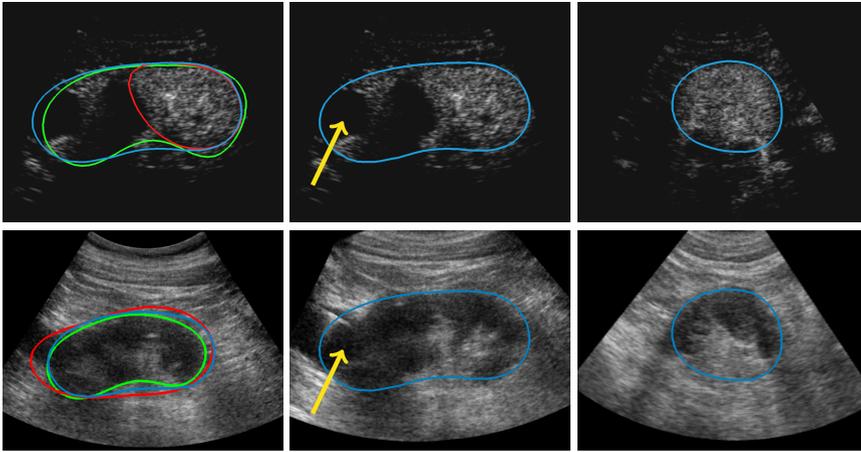


**Fig. 4.** Comparison of classification results for the single decision forest and the auto-context approach. (Left) ROC Curve. (Middle) Precision-Recall curve. (Right) Boxplot of the Dice coefficients ( $p$ -value  $< 10^{-4}$ ).



**Fig. 5.** Boxplots of segmentation results for kidney segmentation in US and CEUS images, in terms of Dice coefficients (a-b) and relative volume error (c-d). The proposed co-segmentation compares favorably to independent segmentation following [13], with a  $p$ -value  $< 10^{-4}$ .

also segmented the kidney following [13] independently in each modality, which is state-of-the-art for CEUS segmentation. Validation was performed by comparing the segmentation result and the ground truth in both US and CEUS images. Dice coefficients and relative error on the measured kidney volume are reported in Figure 5. Using simultaneously the complementary information from US and CEUS images significantly improves the segmentation accuracy in both modalities. More specifically, the median Dice coefficient is increased from 0.74 to 0.81 in CEUS ( $p$ -value  $< 10^{-4}$ ) and 0.73 to 0.78 in US ( $p$ -value  $< 10^{-4}$ ). Furthermore, the proposed approach provides more reliable clinical information as the median error on the kidney volume is almost divided by two in CEUS (29% versus 15%) and in US (25% versus 13%). Figure 6 shows the joint co-segmentation and registration results for one case. Independent segmentation fails in both US and CEUS images because of the kidney lesion (indicated by the yellow arrow), that looks like the background in CEUS but like the kidney in US. Conversely, the proposed co-segmentation manages to overcome this difficulty by combining information from the two modalities. Furthermore, for this example, one can assess the estimated registration by comparing the location of the lesion in the two modalities. Results on another case were also displayed in Figure 1.



**Fig. 6.** Example of joint co-segmentation and registration for a CEUS (top) and a US (bottom) images. (Left) Comparison of independent segmentations (red) and the proposed co-segmentation (blue) with respect to the ground truths (green). (Middle, Right) Two views of the registered volumes that can be assessed by considering the position of the lesion (yellow arrow).

## 5 Conclusion

In this paper, we introduced a novel framework to jointly perform co-segmentation and registration by seeking the same object in different images. This allowed to significantly improve state-of-the-art results of kidney segmentation in CEUS images by simultaneously taking into account complementary information coming from the available US image. A global CEUS/US registration is also available as a side outcome. The full pipeline is automated and computationally efficient for 3D images.

The genericity of our joint co-segmentation and registration framework is three-fold. First, it can be applied to a large class of variational problems, as shown here with ellipsoid detection and model-based segmentation methods. Second, it can be used for any kind of image: in this paper, we used dedicated classifiers for kidney in US/CEUS images but our approach can be plugged on top of any pixelwise classifier. Finally, we presented co-segmentation in two images but generalization to an arbitrary number of images is straightforward. This paves the way for organ tracking application in 3D+T sequences, which we are currently investigating. We also plan to extend the current framework by considering a non global transformation between images in order to cope with deformable organs.

## References

1. Albrecht, T., et al.: Guidelines for the use of contrast agents in ultrasound. *Ultraschall Med.* 25(4), 249–256 (2004)
2. Breiman, L.: Random forests. *Machine Learning* 45(1), 5–32 (2001)
3. Chan, T.F., Vese, L.A.: Active contours without edges. *IEEE TIP* 10(2), 266–277 (2001)

4. Glocker, B., Pauly, O., Konukoglu, E., Criminisi, A.: Joint classification-regression forests for spatially structured multi-object segmentation. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part IV. LNCS, vol. 7575, pp. 870–881. Springer, Heidelberg (2012)
5. Han, D., Bayouth, J., Song, Q., Taurani, A., Sonka, M., Buatti, J., Wu, X.: Globally optimal tumor segmentation in PET-CT images: A graph-based co-segmentation method. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 245–256. Springer, Heidelberg (2011)
6. Kontschieder, P., Bulò, S.R., Criminisi, A., Kohli, P., Pelillo, M., Bischof, H.: Context-sensitive decision forests for object detection. In: Proceedings of NIPS, pp. 440–448 (2012)
7. Lu, C., Duncan, J.S.: A coupled segmentation and registration framework for medical image analysis using robust point matching and active shape model. In: IEEE Workshop on MMBIA, pp. 129–136 (2012)
8. Martin-Fernandez, M., Alberola-Lopez, C.: An approach for contour detection of human kidneys from ultrasound images using Markov random fields and active contours. *MedIA* 9(1), 1–23 (2005)
9. Montillo, A., Shotton, J., Winn, J., Iglesias, J.E., Metaxas, D., Criminisi, A.: Entangled decision forests and their application for semantic segmentation of CT images. In: Székely, G., Hahn, H.K. (eds.) IPMI 2011. LNCS, vol. 6801, pp. 184–196. Springer, Heidelberg (2011)
10. Mory, B., Somphone, O., Prevost, R., Ardon, R.: Real-time 3D image segmentation by user-constrained template deformation. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 561–568. Springer, Heidelberg (2012)
11. Payet, N., Todorovic, S.: 2-Random Forest Random Field. In: Proceedings of NIPS 2010 (2010)
12. Pohl, K.M., Fisher, J., Grimson, W.E.L., Kikinis, R., Wells, W.M.: A Bayesian model for joint segmentation and registration. *NeuroImage* 31(1), 228–239 (2006)
13. Prevost, R., Mory, B., Correas, J.-M., Cohen, L.D., Ardon, R.: Kidney detection and real-time segmentation in 3D contrast-enhanced ultrasound images. In: Proceedings of IEEE ISBI, pp. 1559–1562 (2012)
14. Saddi, K., Chéfd'hotel, C., Rousson, M., Cheriet, F.: Region-based segmentation via non-rigid template matching. In: Proceedings of ICCV, pp. 1–7 (2007)
15. Tu, Z., Bai, X.: Auto-context and its application to high-level vision tasks and 3D brain image segmentation. *IEEE TPAMI* 32(10), 1744–1757 (2010)
16. Vicente, S., Kolmogorov, V., Rother, C.: Cosegmentation revisited: Models and optimization. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part II. LNCS, vol. 6312, pp. 465–479. Springer, Heidelberg (2010)
17. Wang, F., Vemuri, B.C.: Simultaneous registration and segmentation of anatomical structures from brain MRI. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 17–25. Springer, Heidelberg (2005)
18. Wyatt, P.P., Noble, J.A.: MAP MRF joint segmentation and registration. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002, Part I. LNCS, vol. 2488, pp. 580–587. Springer, Heidelberg (2002)
19. Yezzi, A., Zöllei, L., Kapur, T.: A variational framework for integrating segmentation and registration through active contours. *MedIA* 7(2), 171–185 (2003)
20. Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O.M., Das, T., Jena, R., Price, S.J.: Decision Forests for Tissue-Specific Segmentation of High-Grade Gliomas in Multi-channel MR. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 369–376. Springer, Heidelberg (2012)