Joint Co-Segmentation and Registration of Ultrasound Images

Raphael Prevost¹,², Remi Cuingnet¹, Benoit Mory¹, Jean-Michel Correas³, Laurent D. Cohen², and Roberto Ardon¹

¹ Philips Research Medisys, Suresnes, France
² CEREMADE UMR 7534, Universite Paris Dauphine, Paris, France
³ Adult Radiology Department, Necker Hospital, Paris, France

Abstract. Contrast-enhanced ultrasound (CEUS) is a valuable imaging modality as it allows a visualization of the vascularization and complements the anatomical information provided by conventional ultrasound (US). However, in such images classical segmentation algorithms may be hindered by the noise, the limited field of view or shadowing effects. In this paper, we propose to use simultaneously the different information coming from US and CEUS images to address the problem of kidney segmentation. To that end, we develop a generic framework for joint co-segmentation and registration and apply it to an ellipsoid estimation (kidney detection) and a model-based segmentation algorithm (kidney segmentation). Both methods rely on voxel classification maps, that we estimate using random forest in an original approach. This results in a fully automated pipeline for kidney segmentation in US and CEUS that outperforms state-of-the-art techniques on a clinically representative dataset.

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

1 Introduction

1.1 Clinical Setting

Ultrasound (US) imaging is a popular modality as it is cheap, portable and safe. Contrast-enhanced ultrasound (CEUS) consists in acquiring a peculiar ultrasound image after injecting a contrast agent made of gas-filled microbubbles in the patient’s blood. Because those bubbles have a different acoustic response from the tissues, they can be isolated and images showing only the blood flow can be generated. This modality is particularly interesting when the clinician wants to assess the functioning of highly vascularized organs such as kidneys. Yet, analysis of such images can be very challenging and literature on their segmentation is extremely limited.
In [10], Prevost et al. proposed a method to detect and segment kidneys in CEUS images. While they provided an automated pipeline, failures were reported in several cases and user interactions were needed to obtain a satisfying result. Yet because of shadowing effects, pathologies or restricted field of view, parts of the kidney may be completely invisible in the image. In such cases even expert users cannot confidently delineate the true boundary of the organ.

In this paper, we propose a way to improve segmentation in CEUS images by taking another image into account in order to cope with missing information. Indeed in clinical routine each CEUS acquisition is preceded by a conventional US acquisition to locate the kidney, so two images - that show different information - are actually available. However, automated kidney segmentation in 3D US images is also an open issue. Martin-Fernandez et al. [6] 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 [12]. 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 imaging systems, the two images are in an arbitrary referential and need to be aligned. However classical iconic registration methods are not adapted as 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 tackle them simultaneously. Most of them rely on an iconic registration guiding the segmentation (e.g. [13,9,5]). Yet they assume that the segmentation is known in one of the image, 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. [14] 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 [15], 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 CEUS image presented in [10] to a couple of CEUS and US images. To that end, we develop a generic joint co-segmentation and registration framework
Inspired by [15]. 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. They 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 dataset 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 consist 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 $\phi : \Omega \rightarrow \mathbb{R}$. As in any variational approach, $\phi$ is sought as the minimum of an energy $E$. In the following, we will focus on energies of the following generic form

$$E_I(\phi) = \int_\Omega f(\phi(x)) \ r_I(x) \ dx + R(\phi)$$

(1)

where $f$ is a real-valued function and $r_I(x)$ denotes a pointwise score on whether $x$ looks like an interior or exterior voxel in the image $I$. This is a classical setting in which the segmenting implicit function $\phi$ must achieve a trade-off between an image-based term and a regularization term.
We are interested in the case where a couple of images $I_1 : \Omega \to \mathbb{R}$ and $I_2 : \Omega \to \mathbb{R}$ 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:

$$E_{I_1, I_2}(\phi) = \int_{\Omega_1} f(\phi(x)) \left( r_{I_1}(x) + r_{I_2}(x) \right) \, dx + \mathcal{R}(\phi) .$$

Unfortunately, such an assumption rarely holds in medical applications unless the two images are acquired simultaneously (as in PET-CT imaging [3]). 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 $G_r$. The co-segmentation energy thus reads

$$E_{I_1, I_2}(\phi, G_r) = \int_{\Omega_1} f(\phi(x)) \, r_{I_1}(x) \, dx + \int_{\Omega_2} f(\phi \circ G_r(x)) \, r_{I_2}(x) \, dx + \mathcal{R}(\phi) .$$

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

$$E_{I_1, I_2}(\phi, G_r) = \int_{\Omega_1} f(\phi(x)) \left( r_{I_1}(x) + r_{I_2} \circ G_r^{-1}(x) \right) \, dx + \mathcal{R}(\phi) .$$

Minimizing $E_{I_1, I_2}$ can be therefore be interpreted as performing both segmentation (via $\phi$) and rigid registration (via $G_r$) simultaneously. This generalizes a more classical approach (e.g. [1]) where the images are first aligned in a preprocessing step, and then used in the sense of Eq (2).

In the following, we apply this framework to (i) a robust ellipsoid detection [10] and (ii) implicit template deformation [8] to build a completely automated workflow for kidney segmentation in US and CEUS 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 completely justified.

### 2.2 Robust Ellipsoid Co-Detection

Authors of [10] proposed to detect the kidney in CEUS images as an ellipsoid. For that purpose, they developed a variational framework to achieve fast and robust ellipsoid detection.

Any ellipsoid can be implicitly represented by a function $\phi_{c, M} : \Omega \to \mathbb{R}$ such that $\phi_{c, M}(x) = 1 - (x - c)^T M (x - c)$, where $c \in \mathbb{R}^3$ denotes the ellipsoid center and $M$ is a symmetric positive-definite matrix. The ellipsoid interior is then the zero superlevel set of $\phi_{c, M}$. Given a probability map $p : \Omega \to [0, 1]$ defined at each pixel, the detection is sought as the smallest ellipsoid that includes most of the pixels $x$ with high probability $p(x)$. To limit the influence of possible false positives pixels, a weighting function $w : \Omega \to [0, 1]$ acting on $p$ is simultaneously estimated. The variational problem can then be written as
\[
\min_{c,\mathcal{M}, w} E_d(c, \mathcal{M}, w) = -\int_{\Omega} \phi_{c,\mathcal{M}}(x) \, p(x) \, w(x) \, dx \\
+ \mu \left( \int_{\Omega} p(x) \, w(x) \, dx \right) \cdot \log \left( \frac{\text{Vol}(\mathcal{M})}{|\Omega|} \right)
\]

with \( \text{Vol}(\mathcal{M}) = \frac{4\pi}{3} \sqrt{\det \mathcal{M}^{-1}} \) the ellipsoid volume.

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

- with \( f = Id \) and \( r = -pw \) in the image-based term. \( r \) 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 \( \mathcal{R}(\phi_{c,\mathcal{M}}) = \mathcal{R}(\mathcal{M}) = \mu \int_{\Omega} pw \cdot \log \left( \frac{\text{Vol}(\mathcal{M})}{|\Omega|} \right) \) 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. The purpose of the factor \( \int_{\Omega} pw \) is to normalize the contribution of such a term, while \( \mu \) denotes a trade-off parameter. Its optimal value can be computed (\( \frac{1}{4} \) in 2D and \( \frac{1}{5} \) in 3D) by considering the ideal case where \( p \equiv 1 \) (resp. 0) inside (resp. outside) the ellipsoid.

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

\[
E_{cod}(c, \mathcal{M}, w_{1/2}, \mathcal{G}_r) = -\int_{\Omega} \phi(x) \, p_1(x) \, w_1(x) \, dx - \int_{\Omega} \phi \circ \mathcal{G}_r(x) \, p_2(x) \, w_2(x) \, dx \\
+ \mu \left( \int_{\Omega} p_1 w_1 + p_2 w_2 \right) \cdot \log \left( \frac{\text{Vol}(\mathcal{M})}{|\Omega|} \right)
\]

with \( \text{Vol}(\mathcal{M}) = \frac{4\pi}{3} \sqrt{\det \mathcal{M}^{-1}} \) 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 \( c_2 \) in the second image and the rotation matrix \( R \) :

\[
E_{cod}(c_1, c_2, w_{1/2}, R, \mathcal{M}) = -\int_{\Omega} \phi_{c_1,\mathcal{M}}(x) \, p_1(x) \, w_1(x) \, dx \\
- \int_{\Omega} \phi_{c_2, R^T \mathcal{M} R}(x) \, p_2(x) \, w_2(x) \, dx \\
+ \mu \left( \int_{\Omega} p_1 w_1 + p_2 w_2 \right) \cdot \log \left( \frac{\text{Vol}(\mathcal{M})}{|\Omega|} \right)
\]

Minimization of such energy is similar to the classical ellipsoid detection, and is done in an alternate three-step process:
1. The statistical interpretation still holds for the ellipsoids centers and matrix:
minimizers $c_1^*$ and $c_2^*$ are weighted centroids while $\mathcal{M}^*$ is related to the weighted covariance matrix of pixels coming from both images.

2. The unknown $R$ 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 process is summarized in Algorithm 1. Computational efficiency is maintained: closed-form solutions are available (except for $R$) and the algorithm, though iterative, usually converges in very few iterations.

---

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

**Initialization**
\[
\forall x \in \Omega, \quad w_1(x) \leftarrow 1, \quad w_2(x) \leftarrow 1
\]

**Repeat**

// Estimation of centers $c_1$ and $c_2$ and matrix $\mathcal{M}$
\[
c_1 \leftarrow \frac{1}{\int_{\Omega} p_1 w_1} \int_{\Omega} p_1(x) \ w_1(x) \ x \ dx
\]
\[
c_2 \leftarrow \frac{1}{\int_{\Omega} p_2 w_2} \int_{\Omega} p_2(x) \ w_2(x) \ x \ dx
\]
\[
\mathcal{M}^{-1} \leftarrow \frac{2}{\mu \int_{\Omega} p_1 w_1 + \int_{\Omega} p_2 w_2} \left( \int_{\Omega} p_1(x) \ w_1(x) \ (x - c_1) (x - c_1)^T \ dx
\]
\[
+ \int_{\Omega} p_2(x) \ w_2(x) \ R (x - c_2) (x - c_2)^T R^T \ dx \right)
\]

// Update of the rotation matrix $R$

**Repeat**

$R(\Theta \leftarrow R(\Theta - \Delta t \ \nabla_\Theta E_d(\Theta))$)

**Until** convergence;

// Update of the weighting functions $w_1$ and $w_2$ for each $x \in \Omega$

if $(x - c)^T \mathcal{M} (x - c) \leq 1 - \mu \log \left( \frac{\text{Vol}(\mathcal{M})}{|\Omega|} \right)$ then
\[
w_1(x) \leftarrow 1 \ \text{else} \ \ w_1(x) \leftarrow 0
\]

if $(x - c_2)^T R^T \mathcal{M} R (x - c_2) \leq 1 - \mu \log \left( \frac{\text{Vol}(\mathcal{M})}{|\Omega|} \right)$ then
\[
w_2(x) \leftarrow 1 \ \text{else} \ w_2(x) \leftarrow 0
\]

**Until** 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. The simultaneous use of both images allows a great improvement on the ellipse estimations.
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 [11, 8].

Template deformation is a model-based segmentation framework that represents the segmented object as a deformed initial function (called the template). In an implicit setting [8], this segmentation is represented by the zero-level set of a function $\phi : \Omega \to \mathbb{R}$ defined as $\phi = \phi_0 \circ \psi$, where $\phi_0$ is the implicit template and the transformation $\psi : \Omega \to \Omega$ becomes the unknown of the problem. $\psi$ is sought as a minimum of the following energy

$$E_u(\psi) = \int_{\Omega} H(\phi_0 \circ \psi) \cdot r(x) \; dx + R(\psi),$$

(8)

where $H$ is the Heaviside function (i.e. $H(x) = 1$ if $x > 0$, otherwise 0) and $r$ 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 $R$. 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 $u$ such that $\mathcal{L}(x) = x + (u * K_\sigma)(x)$. $K_\sigma$ 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: $\mathcal{R}(\psi) = \mathcal{R}(\mathcal{L}) = \int_{\Omega} \| \mathcal{L} - Id \|^2 = \int_{\Omega} \| u * K_\sigma \|^2$. Penalizing the magnitude of the displacement field prevents the segmentation to deviate too much from the initial shape prior.

It is clear from Eq (8) that implicit template deformation is part of the framework defined in Eq (1) with $f = H$. We can therefore extend it to cosegmentation by considering the following energy:

$$E_{\text{cos}}(\mathcal{L}, G, G_r) = \int_{\Omega} H(\phi_0 \circ \mathcal{L} \circ G) r_1(x) \, dx$$

$$+ \int_{\Omega} H(\phi_0 \circ \mathcal{L} \circ G \circ G_r) r_2(x) \, dx + \frac{\lambda}{2} \| \mathcal{L} - Id \|^2.$$  \hspace{1cm} (9)

In our application, the template $\phi_0$ is defined as the implicit representation of the detected ellipsoid $\phi_{c_1, M}$. $G$ and $\mathcal{L}$ are initially set to the identity while $G_r$ is initialized with the previously estimated registering transformation: $G_r(x) = R(x + c_1 - c_2)$. The energy $E_{\text{cos}}$ is then minimized with respect to the parameters of $G, G_r$, and each component of the vector field $u$, through a classical gradient descent.

3 Learning Kidney Appearance using Random Forests

CEUS = easy probability (intensity of the image)

But extremely difficult to define a probability in US: high variability of appearance in US images, no standardization intensities, saturation (see Figure 3) However one common thing in appearance: 2 structures in the kidney: parenchyma + sinus We will exploit this structural information in our learning strategy (via random forests)

![Fig. 3. Samples of US images from the database showing the extreme variability of kidney appearance. The structure of the kidney (sinus as a bright area surrounded by a darker parenchyma) is however consistent.](image)
Related work:

[7, 4]: entangled forests / context-sensitive (with regression)
[2] joint regression and classification using SDF
[16] tissue specific segmentation

4 Experiments and Validation

Our test database is composed of 60 couples of US and CEUS volumes acquired from 40 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 X seconds on a standard computer (Intel Core i5 2.67 Ghz, 4GB RAM).

5 Conclusion

References