

This is an Accepted Manuscript of an article published by Taylor & Francis in “*Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*” on 29 May 2015, available online in <http://www.tandfonline.com/10.1080/21681163.2015.1029642>.

## ***Fast Semi-Automatic Segmentation of Focal Liver Lesions in Contrast-Enhanced Ultrasound, based on a Probabilistic Model***

Spyridon Bakas<sup>a†</sup>, Katerina Chatzimichail<sup>b</sup>, Gordon Hunter<sup>a</sup>, Bastien Labbé<sup>c</sup>, Paul S. Sidhu<sup>d</sup> and Dimitrios Makris<sup>a</sup>

<sup>a</sup>*Digital Imaging Research Centre, Faculty of Science, Engineering & Computing, Kingston University, Penrhyn Road, Kingston-upon-Thames, London, KT1 2EE, UK*

*{s.bakas, g.hunter, d.makris}@kingston.ac.uk;*

<sup>b</sup>*Evgenidion Hospital, National & Kapodistrian University, Athens, Greece*

*katerina@hcsl.com;*

<sup>c</sup>*Acquisition & Image Processing, Télécom Physique Strasbourg, France*

*bast.labbe@gmail.com;*

<sup>d</sup>*Department of Diagnostic Radiology, King's College Hospital, Denmark Hill, London, UK*

*paulsidhu@nhs.net*

*(Accepted: 10 Mar 2015 - Published online: 29 May 2015)*

Assessment of focal liver lesions (FLLs) in Contrast-Enhanced Ultrasound requires the delineation of the FLL in at least one frame of the acquired data, which is currently performed manually by experienced radiologists. Such a task leads to subjective results, is time-consuming and prone to misinterpretation and human error. This paper describes an attempt to improve this clinical practice by proposing a novel fast two-step method to automate the FLL segmentation, initialised only by a single seed point. Firstly, rectangular force functions are used to improve the accuracy and computational efficiency of an Active Ellipse model for approximating the FLL shape. Then, a novel Probabilistic Boundary Refinement method is used to iteratively classify boundary pixels rapidly. The proposed method allows for faster and easier assessment of FLLs, whilst requiring less interaction, but producing results comparably consistent with manual delineations, and hence increasing the confidence of radiologists when making a diagnosis. Quantitative evaluation based on real clinical data, from two different European countries reflecting true clinical practice, demonstrates the value of the proposed method.

**Keywords:** Segmentation; Active Ellipse; Focal Liver Lesion; Contrast-Enhanced Ultrasound; Probabilistic Classifier

### **1. Introduction**

Primary liver cancer is the second most common cause of death from cancer worldwide, with the annual mortality rate estimated to be nearly 746,000 deaths and annual incidence rate of almost 783,000 new cases diagnosed throughout the world (W.H.O. 2012). 70-90% of primary liver cancer cases are due to a malignant focal liver lesion (FLL), namely Hepatocellular Carcinoma (Blachier et al. 2013), which is also the fifth most common type of cancer worldwide (Llovet et al. 2003). The potential for the early distinction between a malignant and a benign FLL is of significant importance as, in the former case, there is a greater possibility of non-surgical therapeutic intervention and healing, if diagnosed sufficiently early (i.e. when in a premature/small state). This would result in an improvement to patients' care and eventual survival prospects (Lencioni et al. 2007).

---

<sup>†</sup>Corresponding author. Email: s.bakas@gmail.com, Tel.: +44.74.1414.5776

Contrast-Enhanced Ultrasound (CEUS) is recognised as the most cost-efficient imaging solution for distinguishing between benign and malignant FLLs (Westwood et al. 2013), since it is easy to perform, and uses portable and relatively low cost equipment that allows its presence in every clinic and even at the bedside. Furthermore, CEUS has gained acceptance for the detection and characterisation of very small FLLs –i.e. at their very early/premature stages– with diagnostic accuracy exceeding 95% for the evaluation of malignancies (Strobel et al. 2009).

CEUS is based on the medical ultrasound (US) modality and requires the intravenous injection of a contrast medium, which provides a brightness enhancement to the appearance of the blood flow. The apparent changes of the brightness intensity, during a CEUS examination, describe the perfusion dynamics and lead to the distinction between the nature of tissues (Claudon et al. 2013). Visual examples of these apparent changes for a hypo- and a hyper-enhancing FLL are shown in Fig. 1. Note that these examples identify the behaviour of benign and potentially malignant FLLs, respectively.

Radiologists routinely detect, localise, monitor and evaluate FLLs in video recordings of CEUS liver scans manually, through a very time-consuming series of tasks. These include the identification of a reference frame, where the FLL is sufficiently represented and well-distinguished from the remainder of the image plane in order to manually segment it, and eventually classify it as benign or malignant by monitoring the dynamic behaviour of different regions over time, i.e. their relative changes of brightness intensity (Wilson and Burns 2010). Each of these tasks requires a high level of expertise and provides subjective results, which discourage radiologists from using the technique.

The tasks of monitoring and evaluating the dynamic behaviour of an FLL over the duration of a CEUS screening recording have been addressed by different solutions (Rognin et al. 2006; Huang-Wei et al. 2006; Goertz et al. 2010; Ta et al. 2012; Bakas et al. 2012b,a, 2014b). However, all of them assume the existence of prior initialised regions, such as the FLL and a larger area (e.g. the US mask) including both the FLL and the surrounding healthy liver tissue (i.e. the parenchyma) in a reference frame. Furthermore, the selection of such a frame, as well as the automatic segmentation of the US mask have been addressed in (Bakas et al. 2013) and (Bakas et al. 2012b), respectively. However, no practical solution has ever been suggested specifically for segmenting the FLL boundaries in a single frame of a CEUS recording. Thus, an expert is still needed to manually annotate the FLL in a single frame, after following the standard care protocol. This manual procedure leads to subjective results, is time-consuming, and prone to misinterpretation and human error.

Taking into consideration all these factors and the continuously increasing amount of CEUS data obtained and processed by radiologists, automation of the process is highly desirable and should be of significant importance, leading to results less susceptible to human error. After identifying the reference frame in a CEUS video recording, the radiologist is expected to segment the boundaries of the FLL region in this frame, as part of the classification process to identify the FLL as either benign or malignant. Fully automatic segmentation methods are considered inappropriate for the type of data produced using CEUS, as such data has low signal-to-noise ratio and includes severe changes of appearance through the acquired video sequence. Only semi-automatic methods are further considered, as they combine the expert’s knowledge with the computational capability whilst requiring only minimal human interaction, and hence they are considered reliable enough and appropriate for the segmentation problem addressed within this paper.

Current semi-automatic approaches (e.g. active contours (AC) (Kass et al. 1988), active shape models (ASM) (Cootes and Taylor 1992) and level set methods (LS) (Li et al. 2010)) for segmenting different tissues in medical images are impractical, due to either slow convergence caused by the optimisation of a large number of coefficients, or their dependence on a manual initialisation close to the actual tissue boundaries (Marsousi et al. 2010). Both AC and LS can result in a refined segmentation of complex boundaries and are based on the iterative minimisation of a cost function measuring the variation between the regions inside and outside the contour. In addition, ASM can be considered analogous to AC that respect explicit shape constraints. Similarly, an “Active Ellipse” (AE) model was inspired by the paradigm of AC and focussed on segmenting shapes with approximately elliptical boundaries. An AE model was first introduced for medical US in (Marsousi

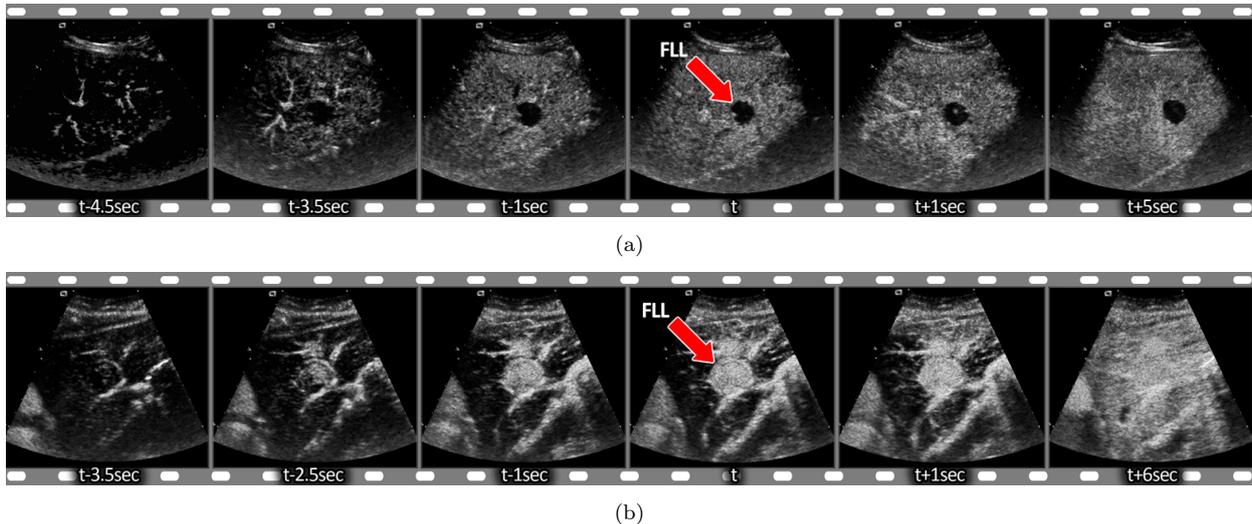


Figure 1. Examples of FLLs. (a) and (b) illustrate the haemodynamic behaviour of a hypo- and a hyper-enhancing FLL, which are the behaviours of benign and potentially malignant FLLs, respectively. The example frames show the behaviour observed in the whole liver from the time that the injected contrast medium reaches the liver (first frame) until the stabilisation of the intensity increase (last frame), including the reference frame (at time  $t$ ), as chosen by the method proposed in Bakas et al. (2013). Note that the appearance of the FLL is essentially the same in frames immediately around the reference frame ( $\pm 1$ sec), but the FLL may be indistinguishable from the parenchyma at the beginning, or at the end, of this sequence.

et al. 2010) to approximate the left ventricle chamber of the human heart. Although no energy function is explicitly defined, as is the case of AC, energy minimisation is implied by the force field used for the iterative evolution of the AE. These forces were based on modified Gaussian functions, the difference of which from the standard Gaussian distribution probability density function (pdf) is illustrated in Fig. 3(b), whilst using the same  $\mu$  and  $\sigma$  values for both functions. However, such smooth functions may not be appropriate to distinguish the FLL boundaries precisely. Furthermore, contraction was not allowed, but only expansion, as these force functions were always positive (Fig. 3(b)).

In this paper, a fast two-step method is proposed for segmenting the FLL in a reference frame of a CEUS video sequence that requires only a single seed point for initialisation, in an attempt to improve the current clinical practice, by assisting radiologists to make a diagnosis more easily and with greater confidence. There are two main technical contributions in our methodology. Firstly, a rectangular force function is used here that improves the AE model proposed in (Marsousi et al. 2010). Our improved AE model requires fewer coefficients for its iterative update and, unlike the approach of (Marsousi et al. 2010), also allows contraction of the AE and faster convergence. Secondly, a novel iterative Probabilistic Boundary Refinement (PBR) method is applied, which outperforms existing iterative approaches, such as Level Sets, in terms of accuracy and computational efficiency. This paper provides extensive evaluation of our method and its preliminary results presented in (Bakas et al. 2014a), through further experiments and more detailed statistical analysis.

## 2. Materials and Methods

### 2.1 Materials

The data used for quantitatively evaluating the proposed method describe real clinical cases from two different European countries with patients in different physical conditions. All data were acquired using Siemens ACUSON US systems (Mountainview, CA). Specifically, 46 cases were captured at King's College Hospital in London (UK), at spatial resolution  $1024 \times 768$  pixels, using an S2000 US system equipped with 4 (or) 6 MHz curvilinear transducer and 14 cases were captured at Evgenidion Hospital in Athens (Greece), at spatial resolution  $768 \times 576$  pixels, using a Sequoia

C512 US system equipped with 6-2 MHz curvilinear transducer. In all examinations, the second generation contrast medium SonoVue (Schneider 1999) (Bracco S.p.A., Italy) was used in a 2.4ml bolus intravenous injection (into an arm vein), which allows excellent depiction of the FLL vascularity and perfusion (Wilson and Burns 2010). Specific acquisition parameters of the equipment for each patient are unknown, as they were set by the radiologist individually at the start of each examination. Examinations were performed by radiologists with 13-16 years of experience in using CEUS. The acquisition method reflects true clinical practice and, together with the combination of data from two different countries, leads to increased variability. Specifically, all data were acquired without prior knowledge of subsequent software processing and without any specific instructions being given to the radiologist beforehand. Appropriate ethics and confidentiality procedures have been followed at all times.

Five of the cases were excluded from the application and evaluation of the proposed method since the FLL was either indistinguishable from the parenchyma, or obliterated in the image plane due to acoustic shadows. Even experienced radiologists would be unable to segment accurately the boundaries of the depicted FLL using just a single frame of such sequences. They would either need to review the whole, or a substantial part, of the CEUS video sequence, or else limit the annotation by using a straight line wherever shadowing occurs. Similarly, for an automated approach to address this issue, a holistic approach should be followed that would consider the whole, or a substantial part, of the acquired sequence and not just a single frame. Therefore, the method proposed here for segmenting an FLL in a reference frame of a CEUS video recording was applied to 55 cases, in which the FLL was well-distinguished, rather than all 60 cases. This proportion (i.e. 5/60 cases, or 8.3%) cannot be considered as standard for this kind of clinical data, and may depend on various factors, such as the physical condition of each patient (e.g. proportion of fat tissue), or how closely acquisition guidelines are followed.

## 2.2 Proposed Method

The proposed method takes as input a frame from an acquired CEUS screening recording, selected automatically as stated in Bakas et al. (2013), and a single seed point, which may be anywhere within the area of the FLL candidate, specified by the radiologist, and provides as output a binary mask depicting the segmented FLL boundaries. Initially, based on the fact that 2D representations of an FLL are approximately elliptical (Westwood et al. 2013), an improved version of a fast AE model is employed to quickly determine the ellipse that best approximates the region of the FLL. The FLL segmentation is then further refined automatically by iteratively classifying boundary pixels rapidly, according to a probabilistic model using the different distributions of the pixel brightness intensities of the FLL and the parenchyma (Par). A visual summary representing the pipeline of this method is shown in Fig. 2.

### 2.2.1 Ellipse Approximation

An improved iterative AE model is proposed here to automatically approximate the elliptical shape of an FLL after being initialised by a single seed point within the FLL region. An elliptical shape model  $e(k, x_c, y_c, r_x, r_y)$  is employed, which may be parametrically defined as:

$$e(k, x_c, y_c, r_x, r_y) = [x(k, x_c, r_x), y(k, y_c, r_y)], k : 1 \leq k \leq K \quad (1)$$

$$x^i(k, x_c^i, r_x^i) = x_c^i + r_x^i \sin(\theta_k), \quad y^i(k, y_c^i, r_y^i) = y_c^i + r_y^i \cos(\theta_k)$$

where  $i$  represents the iteration number,  $x_c^i$  and  $y_c^i$  are the coordinates of the centre of gravity (CoG) of the ellipse at the  $i^{th}$  iteration,  $r_x^i$  and  $r_y^i$  are the lengths of the semi-major axes of the ellipse in the horizontal and vertical directions at the  $i^{th}$  iteration, and  $\theta_k$  is the angle between the

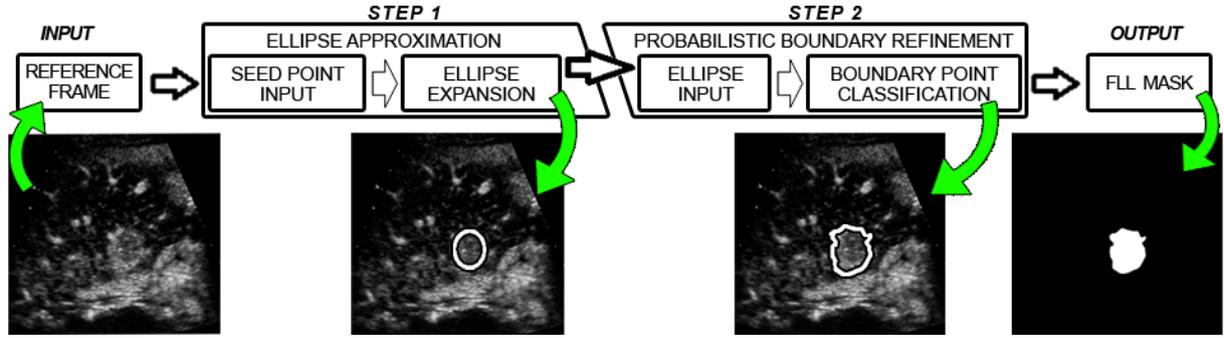


Figure 2. Pipeline of the proposed two-step method for the FLL segmentation. A reference frame from a CEUS recording is used as input to the proposed method, where a single seed point is specified by the radiologist anywhere within the area of the FLL candidate, enabling the first step to approximate the FLL by an ellipse. The second step then refines the boundaries of the ellipse according to a probabilistic model, leading to a binary mask of the segmented FLL as the output of the method.

$k^{th}$  of  $K$  equally-spaced points on the contour of the ellipse and the horizontal axis (Fig. 3(a)).

A Gaussian distribution model with mean  $\mu$  and standard deviation  $\sigma$  is used to describe the distribution of pixel brightness intensities within the AE. Alternative force functions for updating the AE on each iteration are proposed here, for comparison with those proposed in the original AE work (Marsousi et al. 2010). These force functions allow the AE to iteratively expand or contract on each major axis independently. This expansion and/or contraction is performed according to four different forces (Fig. 3(c)). Each of the forces is based on a rectangular function ( $Rect_{2\sigma}(I(p_k))$ ), which is positive for pixel intensities within one standard deviation  $\sigma$  of the mean  $\mu$  of the Gaussian distribution, but negative otherwise (Eq. 2).

$$Rect_{2\sigma}(I(p_k)) = \begin{cases} +1, & \text{if } \mu - \sigma \leq I(p_k) \leq \mu + \sigma \\ -1, & \text{if } I(p_k) < \mu - \sigma, \quad \text{or} \quad I(p_k) > \mu + \sigma \end{cases} \quad (2)$$

where  $I(p_k)$  is the intensity of the pixel  $p_k$  and  $2\sigma$  specifies the width of the interval over which the rectangular function is positive. After considering that the boundary of the AE can be sampled into  $K$  points, then the pixel  $p_k$  is in the  $k^{th}$  of the  $K$  points in this boundary.

The sharpness of the transition of  $Rect_{2\sigma}(I(p_k))$  (Fig. 3(b)) is more appropriate for emphasising the FLL boundaries than the smooth Gaussian pdf (i.e.  $\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ ), the corresponding Gaussian function scaled to have unit peak height (i.e.  $\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ ), and the modified version (i.e.  $\exp\left(-\frac{(x-\mu)^2}{\sigma}\right)$ ) proposed in (Marsousi et al. 2010). Furthermore, the adoption of  $Rect_{2\sigma}(I(p_k))$  allows contraction of the AE, contrary to (Marsousi et al. 2010), and faster convergence. The force proposed here is defined as follows:

$$f_g = \sum_{k=1}^K Rect_{2\sigma}(I(p_k)) \cdot W_g \left(\frac{k}{K}\right), \quad \forall g \in \{x_+, x_-, y_+, y_-\} \quad (3)$$

where  $x$  and  $y$  denote the horizontal and vertical axes for the forces, with the '+' and '-' subscripts depicting the direction of each of the forces (Fig. 3(c)).  $W_g$  are the weightings used with each force for each point of the ellipse, defined as in (Marsousi et al. 2010). Note that the forces are considered positive for an expansion, but negative for a contraction.

At each iteration  $i$ , the parameters of the AE proposed here are updated according to the following formula:

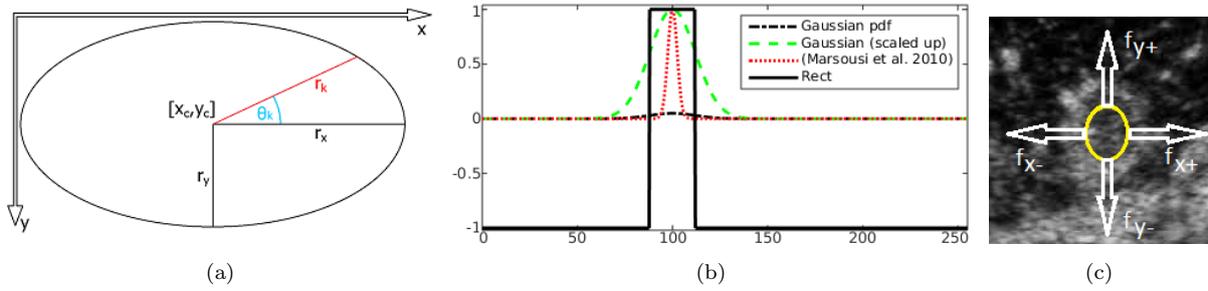


Figure 3. (a) Depiction of the ellipse model ( $e(k, x_c, y_c, r_x, r_y)$ ) used in the proposed method.  $x_c$  and  $y_c$  are the coordinates of the ellipse's CoG,  $r_x$  and  $r_y$  are the lengths of the semi-major axes of the ellipse in the horizontal and vertical directions, and  $\theta_k$  is the angle between the  $k^{th}$  of  $K$  equally-spaced points on the contour of the ellipse and the horizontal axis. (b) Illustration of the difference between the standard Gaussian distribution pdf, the corresponding Gaussian function scaled to have unit peak height, the modified Gaussian function used in (Marsousi et al. 2010), and the rectangular function proposed here for the force functions used for the iterative evolution of the AE. Note that the same  $\mu$  and  $\sigma$  values are used for all four functions, namely  $\mu = 100$  and  $\sigma = 12.5$ . (c) The horizontal and vertical forces applied on the AE proposed here. The forces are considered to be positive for an expansion, but negative for a contraction.

$$\begin{bmatrix} \Delta x_{c,i} \\ \Delta y_{c,i} \\ \Delta r_{x,i} \\ \Delta r_{y,i} \end{bmatrix} = \begin{bmatrix} \zeta & -\zeta & 0 & 0 \\ 0 & 0 & \zeta & -\zeta \\ \zeta & \zeta & 0 & 0 \\ 0 & 0 & \zeta & \zeta \end{bmatrix} \begin{bmatrix} f_{x+} \\ f_{x-} \\ f_{y+} \\ f_{y-} \end{bmatrix} \quad (4)$$

where  $\Delta x_{c,i}$  and  $\Delta y_{c,i}$  represent the changes in the coordinates of the ellipse's CoG ( $x_c, y_c$ ) during that iteration  $i$ ,  $\Delta r_{x,i}$  and  $\Delta r_{y,i}$  are the updates to the semi-major axes during iteration  $i$ ,  $\zeta$  is an ‘‘acceleration parameter’’ and  $f_x, f_y$  are the horizontal and vertical forces applied to the AE, with the plus and minus signs (i.e. for  $f_{x+}, f_{x-}, f_{y+}, f_{y-}$ ) depicting the direction of each of the forces. The updating formula in Eq. 4 omits the constant rate for the evolution of the AE proposed in (Marsousi et al. 2010) resulting to  $\zeta$  being the only one parameter to configure.

Updating the position and shape of the AE terminates when either a maximum number of iterations is reached, or the AE extends outside the boundaries of the workspace, or the value of  $\sum_g |f_g|$  and the changes in the values of  $\mu$  and  $\sigma$  all converge, i.e. each of them stay below a specified tolerance over three successive iterations.

### 2.2.2 Probabilistic Boundary Refinement

This PBR step is initialised by the final ellipse  $e_{AE}(K, x_c, y_c, r_x, r_y)$ , enclosing an area  $A_{AE}$  that resulted from the process described in the previous section. PBR considers a non-parametric generic shape model of a closed contour, sampled using  $K$  points. Initially, another ellipse is created ( $e_{EXT}$ ), concentric with  $e_{AE}$  ( $x_c, y_c$ ) but with semi-major axes double the lengths of those of  $e_{AE}$  (i.e.  $2r_x, 2r_y$ ) and enclosing an area  $A_{EXT}$  (Fig. 4(a)). An annular-shaped area,  $A_{ring}$ , can then be defined as the area between these ellipses, i.e. the set of pixels in  $A_{EXT}$ , but not in  $A_{AE}$  ( $A_{ring} = A_{AE}^C \cap A_{EXT} = A_{EXT} \setminus A_{AE}$ ). The brightness intensity populations of the pixels inside each of the areas  $A_{AE}$  and  $A_{ring}$ , provided that both areas are inside the workspace, are assumed to describe the distributions of the FLL and parenchyma, respectively. However, in some cases, due to the irregular (but still approximately elliptical) apparent FLL shape, some extended protrusions might appear within (or even extend beyond the outer boundary of)  $A_{ring}$ . In such cases, the brightness intensity population of the pixels inside  $A_{ring}$  will include some values from the distribution of the FLL. As shown in the examples of Fig. 8, this does not seem to affect the segmentation results and these irregularities are correctly included in the final FLL segmentation.

Two distinct Gaussian probability density functions are used to model the populations of the pixel intensities inside  $A_{AE}$  and  $A_{ring}$ , respectively (Fig. 4(b) & 4(c)). Specifically, the proposed

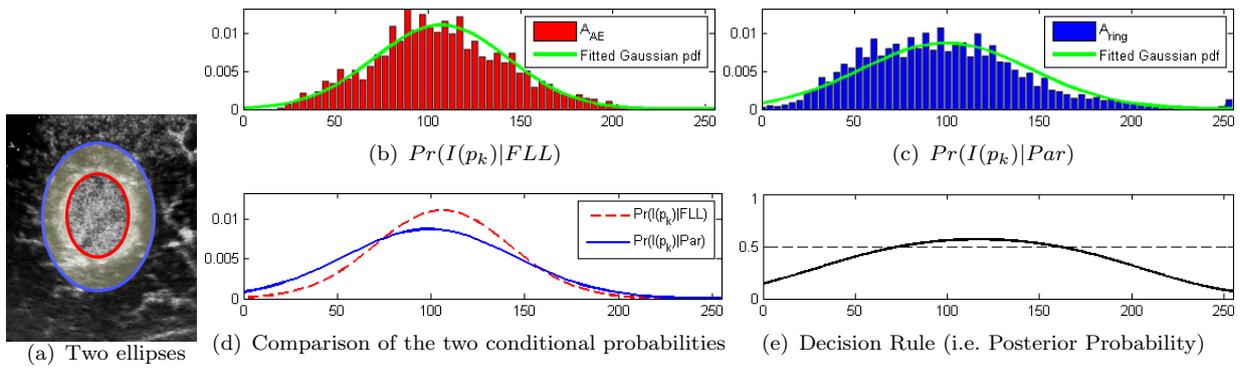


Figure 4. (a) Two concentric ellipses approximating the FLL in an example case. The red (inner) ellipse  $e_{AE}(x_c, y_c, r_x, r_y)$  is the output of the AE model and the blue (outer) ellipse  $e_{EXT}(x_c, y_c, 2r_x, 2r_y)$  is the one employed to enclose the pixels immediately outside  $e_{AE}$ , in a region  $A_{ring}$ . The top two graphs (b & c) illustrate the class conditional probability density of each class, by fitting a Gaussian model to the populations of the pixel intensities using MLE. Graph (d) shows the comparison between the two conditional probability densities. Graph (e) shows the posterior probability for “FLL” given the observed intensity value, after using Bayes’ Theorem with equal priors for the two classes. A pixel is categorised as belonging to “FLL” if this exceeds 0.5, and to “Par” otherwise. Note that the horizontal axis on all graphs shows the possible pixel intensity values.

approach first processes samples from the pixel intensity population of each class separately and applies a normalisation to the histograms of both populations. It then models the class-conditional probability densities ( $Pr(I(p_k)|FLL)$  and  $Pr(I(p_k)|Par)$ ) by fitting a Gaussian model, using Maximum Likelihood Estimation (MLE) to find the mean and standard deviation, for each class.

Consequently, the segmentation of the FLL’s boundary is treated as a binary classification problem, allocating pixels to either “FLL” or “Par” for approximating the boundaries of the FLL with more accuracy. Specifically, at each iteration, the intensity  $I(p_k)$  of each boundary point of the closed contour  $p_k$  is assessed and  $Pr(I(p_k)|FLL)$  is compared with  $Pr(I(p_k)|Par)$  (Fig. 4(d)). This point  $p_k$  is classified into FLL or Par according to the larger of the two conditional probabilities. This is equivalent to a classification based on Bayes’ Theorem with equal priors for the two classes, i.e.  $Pr(FLL) = Pr(Par) = 0.5$  (Fig. 4(e)). The contour is then expanded or contracted along the radial line from the CoG to  $p_k$ , based on this classification decision, so that  $p_k$  is then either inside (FLL) or outside (Par) the contour.

The updating of the shape of the closed contour terminates if either a maximum number of iterations is reached, or the closed contour extends outside the workspace, or the Euclidean length of the closed contour converges over five successive iterations, i.e. the changes in the length over these iterations all stay below a specified tolerance.

### 3. Experiments and Results

#### 3.1 Evaluation Metric

To evaluate quantitatively the proposed method at the pixel level, firstly the ground truth (GT) silhouettes of the FLLs ( $F_{GT}$ ) were manually annotated by a radiologist in the reference frame. These silhouettes are then compared with the areas automatically segmented by the proposed method,  $F_d$  (Fig. 8) using the overlap metric of the Jaccard index as a percentage:  $J = \frac{|F_{GT} \cap F_d|}{|F_{GT} \cup F_d|}$ . The Jaccard index,  $J$ , for an FLL includes information only from pixels  $p$  that belong to the FLL both in the GT and in the system’s decision ( $p \in F_{GT} \cap F_d$ ) but penalises pixels misclassified as either FLL or non-FLL, i.e.  $p \in (F_{GT} \cup F_d) \setminus (F_{GT} \cap F_d)$ .

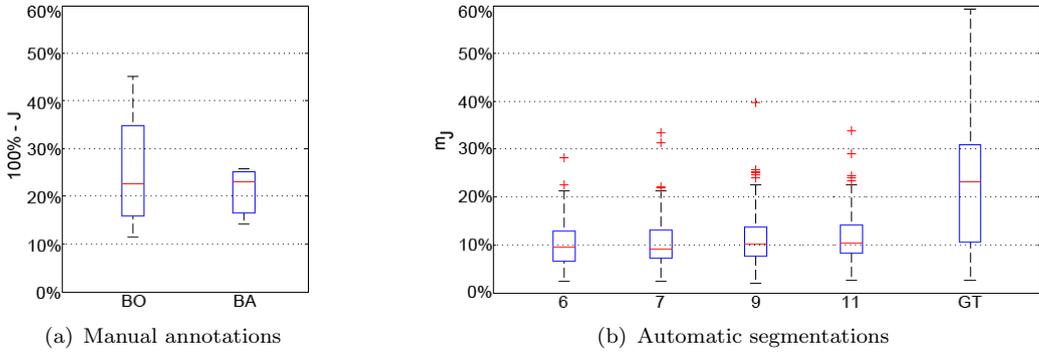


Figure 5. (a) The variation in the disagreement of manual annotations, between operators ‘BO’ (i.e. inter-operator) and between attempts by the same operator ‘BA’ (i.e. intra-operator). (b) The segmentation inconsistency for multiple automatic segmentations across the 55 clinical cases using  $m_J$ , i.e. the median value of the difference between  $J$  and 100% over all segmentations resulting from the 100 different seed points for each case. The first four box-and-whisker plots show results for randomly selected seed points anywhere within a circular area ( $F_{small}$ ) over various values of  $r_F$  (i.e. 6, 7, 9 and 11), and ‘GT’ shows results for seed points randomly selected anywhere within  $F_{GT}$ . Smaller values on both (a) and (b) indicate more consistent segmentation and, particularly for (b), less dependency on the seed point positioning.

## 3.2 Segmentation Consistency

### 3.2.1 Differences between Manual Annotations (Inter- and Intra-operator variation)

Due to manual annotation of  $F_{GT}$ , based on the visual interpretation of the FLL boundaries by the operator, some difference is expected in the  $F_{GT}$  specified by different radiologists (i.e. inter-operator variation – the annotation is non-reproducible), or even by the same radiologist on a different occasion (i.e. intra-operator variation – the annotation is non-repeatable). Based on the degree of this difference, an intrinsic error is introduced, which determines what should be considered by radiologists to be the acceptable margin of “error” during the evaluation of the proposed method.

Specifically, 11 cases were each annotated by 2 different clinicians, and 3 cases were each annotated twice by the same clinician, with a two-week gap between the attempts. The results of the agreement between the different annotations yield the median value of  $J$  to be approximately 76% for both situations of inter- and intra-operator agreement. Therefore, the median value of the inter- and intra-operator variation, across the provided clinical cases, determines what could be considered the acceptable margin of “error” in  $J$  (i.e. the difference of  $J$  from 100%), which is approximately 24% for both situations, namely between (inter-) operators and between attempts by the same operator (intra-), i.e. ‘BO’ and ‘BA’, respectively, in Fig. 5(a). In addition, the inter-quartile ranges (IQR) of the inter- and the intra-operator variation, again in terms of the  $J$  values being expressed as percentages, across the provided clinical cases, are equal to 19.3% and 8.7%, respectively.

### 3.2.2 Dependency on Seed Point Positioning

To evaluate the effect of variation in the manual positioning of the single seed point by the radiologist on the final segmentation, the variation in results due to the use of 100 randomly selected seed points within the whole  $F_{GT}$  was investigated. The value of the parameter  $\zeta$  (Eq. 4) was set equal to 1 in all cases and experiments described in this section.

Furthermore, as the random selection of a point within the whole  $F_{GT}$  might provide seed points that a radiologist would never select (e.g. a seed point on the FLL boundaries), an additional experiment was carried out, where 100 distinct random positions were selected within a circular area ( $F_{small}$ ), of radius  $r_F$ , centred on each of the manually selected seed points initially provided. We avoid setting  $r_F$  dynamically, as a specified fraction of the size of the FLL, because this would be problematic in cases where the FLL is very small, since the resulting circular area would contain less than 100 potential seed points.

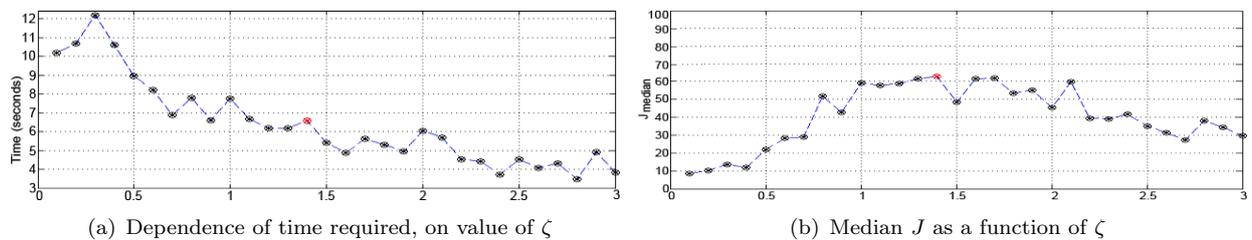


Figure 6. Effect of the acceleration parameter  $\zeta$ . (a) The relationship between the value of parameter  $\zeta$  ( $x$  axis) and the average computational time required for the proposed method to converge ( $y$  axis). It is observed that  $\zeta$  is inversely proportional to the required time. In (b) the  $y$  axis shows the median value of the distribution of the values of  $J$  across all cases for each tested value of  $\zeta$  ( $x$  axis). Note that the value of  $J$  stays relatively stable for  $\zeta \in [1, 1.7]$  and its maximum value is equal to 62.91% when  $\zeta = 1.4$ .

A “median segmentation” ( $MS$ ) is created for each case, by considering the pixels included in more than 50% of the segmentations resulting from each seed point. The difference between  $J$  and 100% is computed when each segmentation result is compared with the  $MS$  for each case. The median value of these differences of  $J$  over all segmentations (i.e.  $m_J$ ) is then used to show the inconsistency in the various segmentations for each individual case, i.e. the variation due to different selections of the seed point. This evaluation was performed for different values of  $r_F$  and the results are shown in Fig. 5(b).

Specifically, for seed points selected within  $F_{small}$  the median values (across the different clinical cases) of  $m_J$ , for  $r_F \in \{6, 7, 9, 11\}$ , are well under the corresponding median values of the disagreement between the various manual annotations for both ‘BO’ and ‘BA’, (i.e. the inter- and intra-operator variation), namely in the range [9.14% – 10.34%]. In addition, the first, the third quartile and the IQR of  $m_J$  all show more consistent segmentation, for all values of  $r_F$  investigated and across the provided clinical cases, when compared with the disagreement between the manual annotations (Fig. 5(a)), and therefore show little dependence on the positioning of the seed point. Specifically, the IQR is equal to 6.38%, 5.32%, 6.38%, and 6.17% for the values of  $r_F$  equal to 6, 7, 9, and 11, respectively. Furthermore, for seed points selected within  $F_{GT}$ , the median value of  $m_J$  is approximately 24% and hence comparable with the median value of the inter- and intra-operator variation (i.e. the disagreement between the manual annotations for both ‘BO’ and ‘BA’), and the IQR of  $m_J$  is equal to 20.12% and hence comparable with the equivalent for ‘BO’ (Fig. 5(a)). This shows that the segmentation results from randomly selected seed points are, at worst, comparably consistent as the results from distinct manual annotations. Note that the intention of the comparison between the automatic segmentation results based on different input seed points and the manual annotations, is to assess only the consistency of the former and not their accuracy.

### 3.3 Effect of the Acceleration Parameter $\zeta$

In addition to investigating the performance of the proposed method for a fixed value of the only configurable parameter  $\zeta$ , identification of the effect of changes in its value is performed using an exhaustive search across the provided clinical cases. This was carried out for values of the acceleration parameter  $\zeta$  in the range  $\{0.1, 3\}$ , with an increment step 0.1. For this approach, the proposed two-step segmentation method was applied to the 55 clinical cases and the resulted FLL boundaries were compared with  $F_{GT}$  in the same frame and values of  $J$  were calculated for each.

The results (Fig. 6(a)) clearly show the dependence between the value of  $\zeta$  and the average computational time required for the method to converge per case. Specifically, as the value of  $\zeta$  increases, the required time generally decreases. Furthermore, results in Fig. 6(b) show the median values of the distribution of  $J$  values, obtained across all the cases with an annotated GT, for each value of  $\zeta$ . Specifically, it is observed that the value of  $J$  stays relatively stable for  $\zeta \in [1, 1.7]$  and its maximum is equal to 62.91% for  $\zeta = 1.4$ .

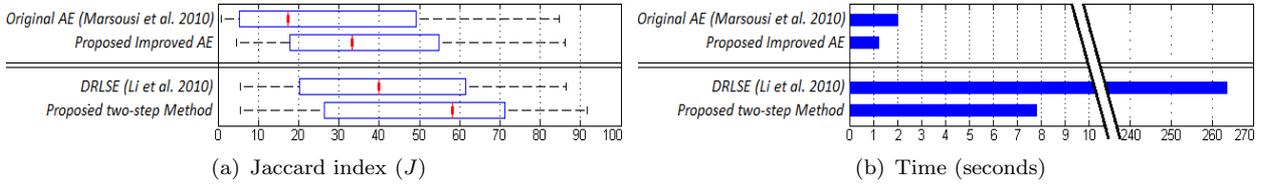


Figure 7. The graph (a) shows  $J$  for assessing the variation of the overlap between  $F_{GT}$  and  $F_d$ , across the provided clinical cases, for the original AE (Marsousi et al. 2010), the proposed improved AE, the LS (DRLSE) method (Li et al. 2010) and the proposed two-step method. Comparison of the median value, the first and the third quartiles, as well as the IQR of  $J$  between the proposed improved AE and the original AE justifies the use of the rectangular force function. Comparison of the same elements between the second step of the proposed method (PBR) and the DRLSE method, both initialised with the output of the proposed AE model, shows that the proposed method provides a generally superior FLL segmentation. Statistical significance of the results is given by the two-tailed p-values, of paired sample t-tests, being less than 0.05. (b) The proposed two-step method converges on average in just 3% of the time required by the DRLSE approach to converge, i.e. 8 seconds versus 264 seconds, across the provided dataset. This, in combination with the superior segmentation performance, would assist in determining the suitability and applicability of the proposed method in real clinical practice.

### 3.4 Comparison to Previous Work

The performance and efficiency of the proposed method are compared against the original AE (Marsousi et al. 2010) and the fastest LS, namely Distance Regularized Level Set Evolution (DRLSE) (Li et al. 2010), both implemented on the same hardware and tested on the same clinical cases. The value of the acceleration parameter  $\zeta$  in Eq. 4 was set equal to 1 for all clinical cases, for both the conventional and proposed AE method.

Comparative results for  $J$  are shown in Fig. 7(a). Comparing the median value and the IQR of  $J$  between the proposed improved AE and the original AE (Marsousi et al. 2010) justifies the choice of the rectangular force function. Application of the proposed PBR step to the output ( $e_{AE}$ ) from the AE gives further improvement to  $J$ , indicating a much better segmentation of the FLL. In addition, the results from the application of the PBR are compared with results of the DRLSE (Li et al. 2010) method, again initialised by  $e_{AE}$ . Both the median value and the upper and lower quartiles of  $J$  across all cases are higher for PBR than for DRLSE, showing that the FLL segmentation is generally superior when using the proposed method. The results shown in Fig. 7(a) are statistically significant, as the paired sample t-tests performed for both pairs of methods yielded two-tailed p-values smaller than 0.05. Specifically, for the method pair of the proposed AE and the original AE, the two-tailed p-value is 0.0146, and for the method pair of the PBR and the DRLSE,  $p < 0.00012$ . Use of the paired sample t-tests is justified in this situation, since the results from each method are compared on exactly the same set of case studies. Furthermore, the proposed method tends to undersegment the FLL, when compared with  $F_{GT}$ , instead of oversegmenting it (Fig. 8). This undersegmentation, by the complete pipeline of our proposed two-step method, might occur due to the undersegmentation performed by the first step (AE) of the pipeline. The AE segments the FLL conservatively, as it stops growing (or even contracts) once the intensity of the pixels immediately outside the AE differs from the mean intensity within the evolving AE. Therefore, a part of the FLL tends to be included within  $A_{ring}$  when modelling the intensity distributions during the second step (namely the PBR) of the methodology.

The suitability and applicability of the applied methods in real clinical practice, in addition to the segmentation performance, are dependent on the computational time that each of the methods requires. For comparison purposes, the methods described above were applied to all the clinical cases provided and the time required for each was averaged across those cases. This shows that the enhanced AE method needs on average about only 60% of the time required for the AE of (Marsousi et al. 2010) to converge, i.e. 1.2 rather than 2 seconds per case, as the rectangular force functions lead to faster ellipse fitting. In addition, the proposed two-step method on average converges in a very small fraction (2.2%) of the time required by DRLSE, e.g. approximately 8 seconds rather than 264 seconds (4.4 minutes) per case (Fig. 7(b)).

According to the results obtained, the proposed two-step method may be considered appropriate for use in real clinical practice, due to both its much reduced demands in terms of computational

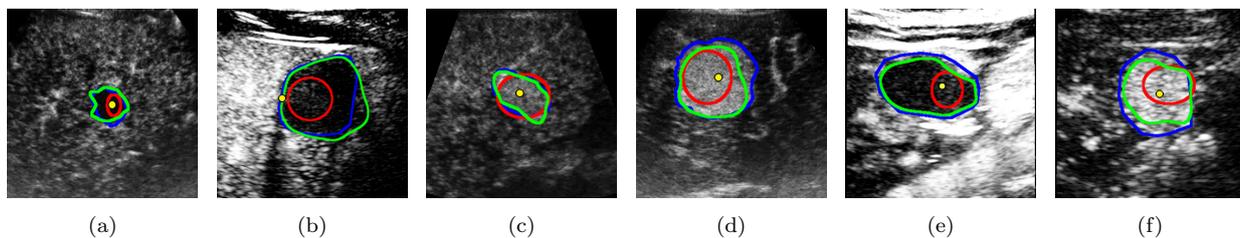


Figure 8. Example output results of the first step (AE) and the final segmentation of the proposed method in comparison with the ground truth ( $F_{GT}$ ) manually annotated by a radiologist. All contours and seed points are dilated for visualisation purposes. The blue contours denote  $F_{GT}$ , the yellow dots denote the seed points provided, the red contours denote the approximation by the enhanced AE model (step 1 -  $e_{AE}$ ) and the green contours denote the final segmentation results of the proposed two-step method. The proposed method clearly gives better approximation to  $F_{GT}$  than does the AE alone - not only by expanding the AE approximation but also by contracting it towards the FLL boundaries, as in (c) and (f). The final segmentation considers irregularities of the apparent FLL shape that were initially included within  $A_{ring}$ , and even in cases that some part of the FLL extended beyond the outer boundary of  $A_{ring}$ , as in (a), (b) and (e). Further observation shows that the proposed method mostly undersegments the FLL. The only example of oversegmentation is shown in (b) where the radiologist has limited  $F_{GT}$  by drawing a straight line, as the FLL boundaries on its right side are not clearly defined due to a shadowing effect.

time and its improved segmentation performance when compared to the other segmentation methods. The clinically acceptable level of results is related to the individual clinical application. For example, if the segmentation result of the proposed method is used to assess the haemodynamic behaviour of the FLL during an acquired CEUS video, then retrieving intensity information from a much smaller area within the FLL (i.e. an undersegmented FLL) is sufficient, and also preferable to including intensity information from the parenchyma region (i.e. oversegmented FLL), as long as the FLL has not been treated/ablated. If the clinical application is to assess the effectiveness of a treatment (e.g. ablation), which requires assessment of the explicit FLL boundaries, then the acceptable level of results should be more strict. Qualitative example results from the proposed method are shown in Fig. 8.

#### 4. Conclusions

In this paper, a novel fast two-step semi-automatic method was proposed for segmenting FLL boundaries in CEUS images, based on the sole manual input of a single seed point, assuming that the FLL is visually distinguishable from the parenchyma and not obliterated in the image plane due to acoustic shadows. The first step of this method consists of employing an improved AE model to approximate the FLL by an ellipse. Specifically, a rectangular force function is proposed which allows contraction of the AE, unlike the original method (Marsousi et al. 2010), and faster convergence. The second step, Probabilistic Boundary Refinement, is used for refining the lesion segmentation by iteratively classifying boundary pixels rapidly, according to a probabilistic model. Furthermore, the proposed two-step method demonstrates better performance, in terms of both the segmentation results and lower computational time requirements, when compared with the fastest LS method (Li et al. 2010).

The proposed method leads to easier and faster assessment of FLLs, as it requires less interaction by radiologists than the explicit manual delineation of an FLL, whilst being comparably consistent, and therefore should increase the confidence of radiologists when making a diagnosis. The proposed method may also be appropriate to other modalities and application areas, e.g. breast US. The limitation of the currently proposed method for the particular problem is the use of a single frame of a CEUS video sequence, which has been shown to be problematic in cases where shadowing effects occur. To address this issue, the current approach should consider the whole, or at least a substantial part around the input frame, of the acquired video sequence, thus being transformed into a more holistic approach, in order to offer a better segmentation of the FLL boundaries that are not apparent in the current input frame. Finally, coupling of the proposed pipeline with an existing tracking method (Rognin et al. 2006; Bakas et al. 2012b; Ta et al. 2012; Bakas et al.

2012a, 2014b), could lead to an automated framework for FLL evaluation, providing an assistive second-opinion tool to radiologists.

## References

- S. Bakas, K. Chatzimichail, A. Hoppe, V. Galariotis, G. Hunter, and D. Makris. Histogram-based motion segmentation & characterisation of focal liver lesions in CEUS. *Annals of the British Machine Vision Association*, 2012(7):1–14, 2012a.
- S. Bakas, A. Hoppe, K. Chatzimichail, V. Galariotis, G. Hunter, and D. Makris. Focal liver lesion tracking in ceus for characterisation based on dynamic behaviour. *Advances in Visual Computing, Springer Berlin Heidelberg, Lecture Notes in Computer Science*, 7431:32–41, 2012b. URL {[http://www.dx.doi.org/10.1007/978-3-642-33179-4\\_4](http://www.dx.doi.org/10.1007/978-3-642-33179-4_4)}.
- S. Bakas, G. Hunter, D. Makris, and C. Thiebaud. Spot the best frame: Towards intelligent automated selection of the optimal frame for initialisation of focal liver lesion candidates in contrast-enhanced ultrasound video sequences. *In Proceedings of the 9th International Conference on Intelligent Environments, Athens, Greece. IEEE*, pages 196–203, 2013. URL {<http://www.dx.doi.org/10.1109/IE.2013.20>}.
- S. Bakas, B. Labbé, G. J. A. Hunter, P. S. Sidhu, K. Chatzimichail, and D. Makris. Fast segmentation of focal liver lesions in contrast-enhanced ultrasound data. *In Proceedings of Medical Image Understanding and Analysis, London, UK. City University*, pages 73–78, 2014a. URL {<http://www.dx.doi.org/10.13140/2.1.1769.0400>}.
- S. Bakas, P. S. Sidhu, M. E. Sellars, G. J. A. Hunter, D. Makris, and K. Chatzimichail. Non-invasive offline characterisation of contrast-enhanced ultrasound evaluations of focal liver lesions: Dynamic assessment using a new tracking method. *20th European Congress of Radiology, Vienna, Austria*, 2014b. URL {<http://www.dx.doi.org/10.1594/ecr2014/C-1378>}.
- M. Blachier, H. Leleu, M. Peck-Radosavljevic, D. C. Valla, and F. Roudot-Thoraval. The burden of liver disease in Europe: A review of available epidemiological data. *European Association for the Study of the Liver*, 2013. URL {<http://www.dx.doi.org/10.1016/j.jhep.2012.12.005>}.
- M. Claudon, C. F. Dietrich, B. I. Choi, D. O. Cosgrove, M. Kudo, C. P. Nolsoe, and et al. Guidelines and Good Clinical Practice Recommendations for CEUS in the Liver - Update 2012: A WFUMB-EFSUMB Initiative in Cooperation with Representatives of AFSUMB, AIUM, ASUM, FLAUS and ICUS. *Ultrasound in Medicine & Biology*, 39(2):187–210, 2013. URL {<http://www.dx.doi.org/10.1016/j.ultrasmedbio.2012.09.002>}.
- T. F. Cootes and C. J. Taylor. Active shape models-‘smart snakes’. *British Machine Vision Conference*, pages 266–275, 1992.
- R. S. Goertz, T. Bernatik, D. Strobel, E. G. Hahn, and T. Haendl. Software-based quantification of contrast-enhanced ultrasound focal liver lesions-a feasibility study. *European Journal of Radiology*, 75(2):e22–e26, 2010. URL {<http://www.dx.doi.org/10.1016/j.ejrad.2009.11.004>}.
- C. Huang-Wei, A. Bleuzen, P. Bourlier, J. Roumy, A. Bouakaz, and L. Pourcelot et al. Differential diagnosis of focal nodular hyperplasia with quantitative parametric analysis in contrast-enhanced sonography. *Investigative Radiology*, 41(3):363–368, 2006. URL {<http://www.dx.doi.org/10.1097/O1.rli.0000195835.56589.55>}.
- M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988. URL {<http://www.dx.doi.org/10.1007/BF00133570>}.
- R. Lencioni, C. D. Pina, L. Crocetti, E. Bozzi, and D. Cioni. Clinical management of focal liver lesions: the key role of real-time contrast-enhanced US. *European Radiology Supplements*, 17(S6):73–79, 2007. URL {<http://www.dx.doi.org/10.1007/s10406-007-0231-8>}.
- C. Li, C. Xu, C. Gui, and M. D. Fox. Distance regularized level set evolution and its application to image segmentation. *IEEE Transactions on Image Processing*, 19(12):3243–3254, 2010. URL {<http://www.dx.doi.org/10.1109/TIP.2010.2069690>}.
- J. M. Llovet, A. Burroughs, and J. Bruix. Hepatocellular carcinoma. *Lancet*, 362:1907–1917, 2003. URL {[http://www.dx.doi.org/10.1016/S0140-6736\(03\)14964-1](http://www.dx.doi.org/10.1016/S0140-6736(03)14964-1)}.
- M. Marsousi, J. Alirezaie, A. Ahmadian, and A. Kocharian. Segmenting echocardiography images using B-spline snake and active ellipse model. *Engineering in Medicine & Biology Society, Annual International Conference of the IEEE*, pages 3125–3128, 2010.
- N. Rognin, R. Campos, J. P. Thiran, T. Messager, A. Broillet, and P. Frinking et al. A new approach

- for automatic motion compensation for improved estimation of perfusion quantification parameters in ultrasound imaging. *In Proceedings of the Eighth French Conference on Acoustics (France)*, pages 61–65, 2006.
- M. Schneider. Characteristics of SonoVue. *Echocardiography*, 16(s1):743–746, 1999. URL <http://www.doi.org/10.1111/j.1540-8175.1999.tb00144.x>.
- D. Strobel, K. Seitz, W. Blank, A. Schuler, C. F. Dietrich, and A. von Herbay et al. Tumor-specific vascularization pattern of liver metastasis, hepatocellular carcinoma, hemangioma and focal nodular hyperplasia in the differential diagnosis of 1349 liver lesions in contrast-enhanced ultrasound (CEUS). *Ultraschall in der Medizin*, 30(4):376–382, 2009. URL <http://www.doi.org/10.1055/s-0028-1109672>.
- C. N. Ta, Y. Kono, C. V. Barback, R. F. Mattrey, and A. C. Kummel. Automating tumor classification with pixel-by-pixel contrast-enhanced ultrasound perfusion kinetics. *Journal of Vacuum Science & Technology B*, 30(2):02C103, 2012.
- M. E. Westwood, M. A. Joore, J. P. C. Grutters, W. K. Redekop, N. Armstrong, and K. Lee et al. Contrast-enhanced ultrasound using SonoVue®(Sulphur Hexafluoride microbubbles) compared with contrast-enhanced computed tomography and contrast-enhanced magnetic resonance imaging for the characterisation of focal liver lesions and detection of liver metastases: a systematic review and cost-effectiveness analysis. *Health Technology Assessment*, 17(16), 2013. URL <http://www.doi.org/10.3310/hta17160>.
- World Health Organisation W.H.O. Fact sheets by population - GLOBOCAN 2012 : estimated cancer incidence, mortality and prevalence worldwide in 2012. [http://globocan.iarc.fr/Pages/fact\\_sheets\\_cancer.aspx](http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx), 2012. (Online; Last accessed: 23/07/2014).
- S. R. Wilson and P. N. Burns. Microbubble-enhanced US in body imaging. *Radiology*, 257(1):24–39, 2010. URL <http://www.doi.org/10.1148/radiol.10091210>.