Automatic Identification of the Optimal Reference Frame for Segmentation and Quantification of Focal Liver Lesions in Contrast-Enhanced Ultrasound

Authors:
Spyridon Bakas[a,b],
Dimitrios Makris[a],
Gordon J.A. Hunter[a],
Cheng Fang[c],
Paul S. Sidhu[c],
Katerina Chatzimichail[d]

[a] *Digital Information Research Centre (DIRC), School of Computer Science & Mathematics, Faculty of Science, Engineering and Computing (SEC), Kingston University, Penrhyn Road, Kingston-Upon-Thames, London, KT1 2EE, United Kingdom*
[b] *Center for Biomedical Image Computing and Analytics (CBICA), Perelman School of Medicine, University of Pennsylvania, Richards Medical Research Laboratories, Floor 7, 3700 Hamilton Walk, Philadelphia, PA, 19104, USA*
[c] *Department of Radiology, King's College Hospital, Denmark Hill, London, SE5 9RS, United Kingdom*
[d] *Radiology & Imaging Research Centre, Evgenidion Hospital, National and Kapodistrian University, Papadiamantopoulou Street 20, T.K. 115 28, Ilisia, Athens, Greece*

Corresponding Author:
Spyridon Bakas, Ph.D.
Center for Biomedical Image Computing and Analytics (CBICA)
Perelman School of Medicine
University of Pennsylvania
Richards Medical Research Laboratories, Floor 7
3700 Hamilton Walk
Philadelphia
PA, 19104
USA
e-mail: s.bakas@kingston.ac.uk, s.bakas@uphs.upenn.edu, s.bakas@gmail.com
tel: +1.267.752.5209, +44.74.14.14.5776

Abstract

Post-examination interpretation of Contrast-enhanced Ultrasound (CEUS) cine-loops of focal liver lesions (FLLs) requires offline manual assessment by experienced radiologists, which is time-consuming and generates subjective results. Such assessment usually starts by manually identifying a reference frame, where FLL and healthy parenchyma are well-distinguished. This study proposes an automatic computational method to objectively identify the optimal reference frame for distinguishing and hence delineating an FLL, by statistically analysing the temporal intensity variation across the spatially-discretized ultrasonographic image. Level of confidence and clinical value of the proposed method were quantitatively evaluated on retrospective multi-institutional data (n=64) and compared with expert interpretations. Results support the proposed method for facilitating easier, faster and reproducible assessment of FLLs, further increasing the radiologists' confidence in diagnostic decisions. Finally, our method yields a useful training tool for radiologists, widening CEUS use in non-specialist centres, potentially leading to reduced turnaround times, lower patient anxiety and healthcare costs.

Introduction

Contrast-enhanced US (CEUS) is an extension of the basic US technique requiring an intravenous injection of a contrast-enhancing agent that provides enhancement of the blood pool and allows for the distinction of a focal liver lesion (FLL) from the surrounding liver parenchyma, by increasing the contrast between them (Chiorean, et al. 2015, Harvey, et al. 2001, Mischi, et al. 2014, Quaia 2007). CEUS performed immediately following a B-mode US examination often allows immediate diagnosis and appropriate patient management. CEUS has gained acceptance for use in the detection and characterisation of FLLs (Dietrich, et al. 2006, Llovet 2005, Llovet, et al. 2003, Quaia 2011, Quaia 2012, Sidhu 2014, Sporea, et al. 2010, Wilson and Burns 2010), with a comparable sensitivity and specificity to CT and MR imaging (Claudon, et al. 2013). Recently, CEUS using Lumason/SonoVue™ has been approved for diagnostic liver imaging in the United States (Seitz and Strobel 2016). The diagnostic accuracy of CEUS for the evaluation of malignant FLLs is in excess of 95% (Strobel, et al. 2009, Westwood, et al. 2013). Furthermore, CEUS is recognised as the most cost-efficient imaging solution for distinguishing between benign and malignant FLLs (Claudon, et al. 2013, Westwood, et al. 2013), using portable and relatively low cost equipment (Sirli, et al. 2010) that allows it to be available widely and even at the patient's bedside.

Despite these advantages, post-examination interpretation of the CEUS cine-loops is a time-consuming and labour-intensive process, based on manual assessment of the acquired data by highly-experienced and specially-trained radiologists. However, this assessment leads to subjective results (Claudon, et al. 2013) and depends on the clinician's experience when acquiring the data, independent of the patient's physiological status and breathing patterns. Initially, the radiologist reviews the entire recorded sequence to identify a reference frame manually, by visual observation, where the FLL is sufficiently well-distinguished from the parenchyma and well-defined within the US image, which we refer to as the "workspace" (Fig.1). Ideally, this frame is expected to demonstrate maximum contrast between the two regions of interest (ROIs), i.e. the FLL and the parenchyma, to facilitate their differentiation, and normally occurs during the initial "arterial" phase of the examination. Due to the variety of different vascular enhancement patterns of FLLs (Albrecht, et al. 2004, Anaye, et al. 2011, Brannigan, et al. 2004, Claudon, et al. 2013, Nicolau, et al. 2006, Sugimoto, et al. 2009, Wilson and Burns 2006), it is not straightforward to specify a fixed time window to robustly predict the point in time at which the maximum contrast will be achieved for any particular example case, either as an actual time interval, or as a constant fraction of the whole duration of the examination.

The aim of this study is to automatically and deterministically detect the optimal reference frame in a CEUS FLL screening recording where the two ROIs are best distinguishable, hence to assist the radiologists and allow for the segmentation and eventually the quantification of these ROIs. However, neither of these ROIs are localised prior to identifying the reference frame, therefore direct measurement of the contrast between them is impossible. The automatic computational approach proposed here investigates how the variation of intensity across the workspace changes over time, after spatially discretising the workspace into local patches, in an attempt to imitate the manually performed procedure and bring the benefits of computational analytics closer to medical experts and their anatomical knowledge. A CEUS FLL screening recording is assumed to be captured in CEUS mode, after the injection of the contrast agent and with the transducer focused on the liver area.
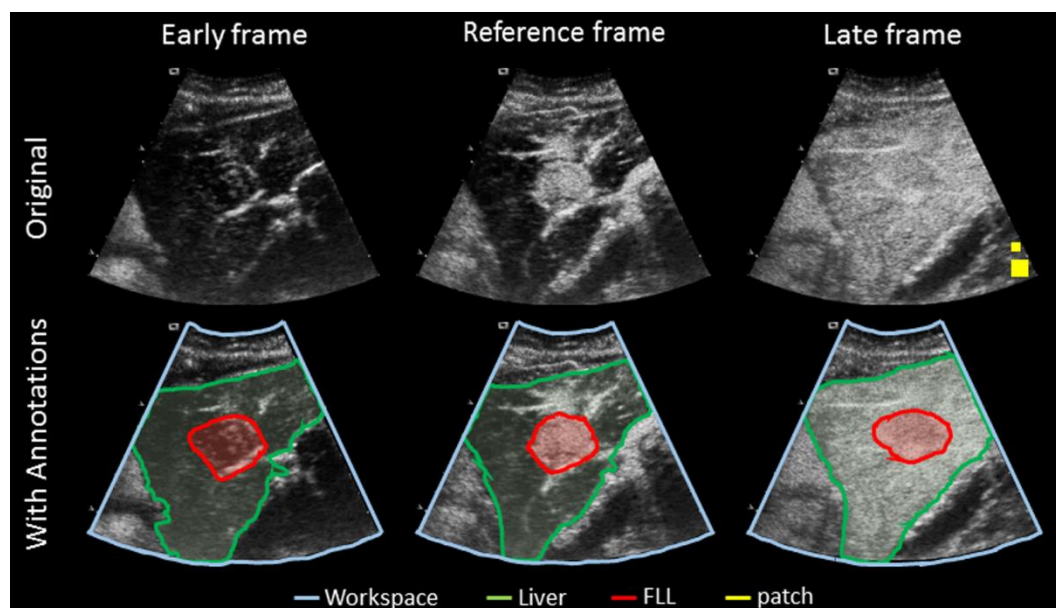
Figure 1. Example frames for the same clinical Contrast-Enhanced Ultrasound cine-loop. The top row of images illustrate the original frames, whereas the bottom row of images have included annotations attempting to approximate the boundaries of the lesion and the surrounding parenchyma, whilst accounting for visible deformations in the image plane. The "early frame" does not depict the lesion or its boundaries at all clearly. The "reference frame" is the radiologist's decision on the frame for the focal liver lesion initialisation, where it is sufficiently well-distinguished from the normal parenchyma and well-defined in the image plane. In the "late frame", the area surrounding the lesion has become isoechoic to the lesion and so is no longer well-defined. The two yellow patches shown in the "Original" "Late frame" denote examples of 32×32 and 16×16 pixels in size.

Materials and Methods

Materials

Data

The clinical data used for evaluating the proposed method, describe retrospectively analysed cohorts obtained with different protocols from two independent institutions; Evgenidion Hospital in Athens (Greece) (Cohort A) and King's College Hospital in London (United Kingdom) (Cohort B), and were treated as a single dataset. The two cohorts comprise a variety of 64 FLLs (Cohort A n=14, Cohort B n=50), median patient age 50 years (range: 19-91 years).

All data included in this study reflect true clinical practice of different protocols, obtained without any prior knowledge of subsequent processing by computer-aided assessment tools and without any specific conditions being instructed to the radiologists beforehand. Therefore, each case consists of a digital CEUS cine-loop covering the complete FLL examination, which in some cases included frames acquired before the transducer was focused on the FLL or even before the contrast agent was injected, or appeared in the image plane. Since the scope of this study is to assess a CEUS FLL screening recording, as defined above (i.e., after the injection of the contrast agent and with the transducer focused on the liver area) we discarded such frames, from the beginning of some sequences, as irrelevant. This resulted in truncated cine-loops with median duration of 24.35 seconds (range: 5.96-19.64 seconds for Cohort A and 3.1-89.83 seconds for Cohort B) including both the arterial and the portal-venous phases. Therefore, in this study, the "first frame" refers to the specified initial frame after all irrelevant frames have been excluded. An example of how such irrelevant information can affect the average brightness intensity of the "workspace" over

time is illustrated in Fig.2 (a). The spatial resolution of the videos for Cohort A is 768×576 pixels, while for Cohort B it is 1024×768 pixels. Examples of benign (n=34) and malignant (n=30) FLLs were included in both cohorts. Local Ethics Board approval waivers were attained at King's College Hospital (UK) and Evgenidion Hospital (GR), where maintenance of randomisation codes for each clinical case and hence confidentiality of patients' records was secured by anonymising all the screened data. Informed consent for using the recordings and diagnostic results for research purposes was acquired from each patient prior to acquiring the scans.

Each CEUS examination/acquisition in Cohort A was performed by a single radiologist with fifteen years of experience using CEUS, and in Cohort B by one of two experienced radiologists with nineteen and fourteen years of experience of use of CEUS, respectively. Only a single lesion was targeted and examined in each patient, even if multiple lesions were identified. Lesion size varied in diameter between 0.5 and 8.0 cm. The physical condition of the patients (e.g. body habitus), was assumed to vary, which increases the intrinsic variability of the data provided and rigour of testing the proposed method.
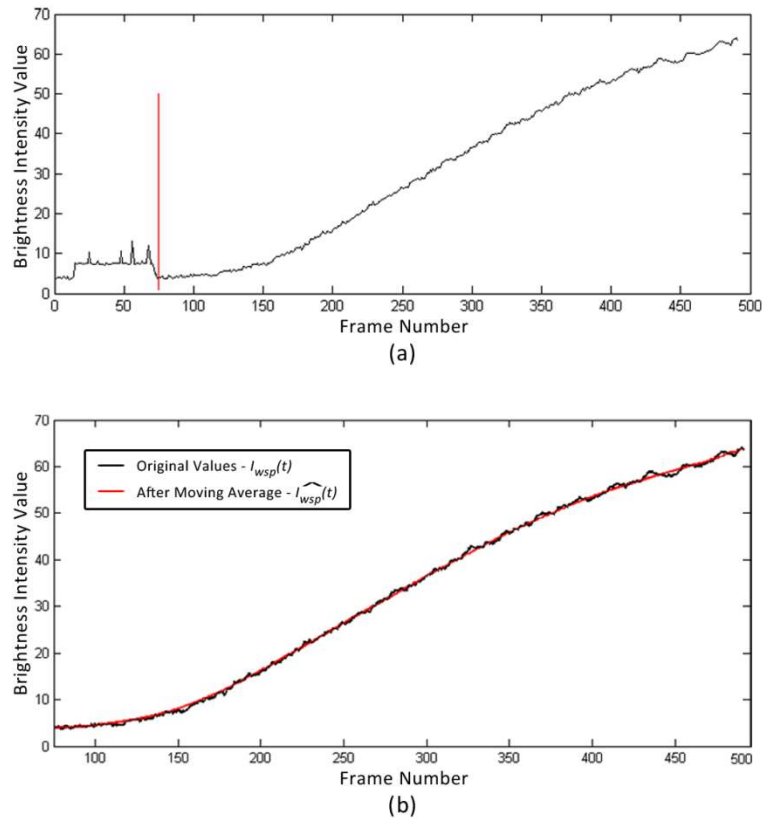

Figure 2. Average brightness intensity of the whole workspace over time (in frames) of an example case from Cohort A. In (a) irrelevant information is included at the beginning of the sequence, namely data captured before the transducer was focused on the region of interest, and this shows up as fluctuations up to the red vertical line (Frame 75), inconsistent with the rest of the graph. In this particular example, the subset of frames of the Contrast-Enhanced Ultrasound cine-loop considered in the subsequent analysis should exclude those first 75 frames. In (b) the black curve shows the original noisy values ($I_{wsp}(t)$) and the red curve those after a `moving average' smoothing has been performed $\widehat{I_{wsp}}(t)$. Note that $\widehat{I_{wsp}}(t)$ is monotonically increasing, in contrast to $I_{wsp}(t)$, which includes many local minima and maxima.

Ground Truth

To evaluate the proposed method quantitatively, and to assess the level of associated confidence in the results, the silhouettes of the FLL and the apparent liver were manually annotated in each frame by a radiologist, providing the Ground Truth (GT) as $F_{GT}(t)$ and $L_{GT}(t)$, respectively, for every frame t. This annotation was based on the perceived local

brightness intensity and shape information varying over the duration of the recordings and the radiologist's expert knowledge. The GT of the parenchyma $P_{GT}(t)$ was then obtained, which is defined as the relative complement of $F_{GT}(t)$ in $L_{GT}(t)$, also known as the set-theoretic difference of $L_{GT}(t)$ and $F_{GT}(t)$, i.e. the set of pixels in $L_{GT}(t)$, but not in $F_{GT}(t)$. This is mathematically defined as: $P_{GT}(t) = F_{GT}(t)^C \cap L_{GT}(t) = L_{GT}(t) \backslash F_{GT}(t)$, where $F_{GT}(t)$ and $P_{GT}(t)$ are the specified subsets of $L_{GT}(t)$.

Equipment

Siemens ACUSON US Systems (Mountain View, CA) were used for the acquisition of all data. Specifically, data on Cohort A were captured using a 6-2 MHz curvilinear transducer on a Sequoia C512, whilst Cohort B data were collected using a 4 or 6 MHz curvilinear transducer on an S2000. All cases in Cohort A were captured at the same frame rate of 25 frames per second (fps), whereas the cases of Cohort B were captured at different frame rates (between 6 and 25 fps). The examinations in both cohorts were performed employing a low mechanical index (Ding, et al. 2005) – at or below 0.10 – technique (Cadence Contrast Pulse Sequencing, CPS, Siemens, Mountain View, CA) (Claudon, et al. 2013, Skyba, et al. 1998). Additional specific acquisition parameters (e.g. field depth, acoustic gain) of the equipment for each patient were variable, as they were set by each radiologist individually at the start of each examination. This reflects true clinical practice, leading to increased variability within the cohorts. In addition, each cine-loop's acquisition was dependent on each operator's preferences during the clinical examination, also contributing to the variability.

A 2.4 ml bolus intravenous injection of SonoVue™ (Bracco S.p.A., Milan, Italy) (Schneider 1999) was used in all the examinations.

Spatiotemporal Intensity Variation (SIV) Method

The method proposed here, Spatiotemporal Intensity Variation (SIV), for identifying the optimal reference frame, takes a CEUS cine-loop sequence as input and automatically identifies a single reference frame as its output.

Considering that, according to radiologists, the optimal reference frame is ideally the one with the maximum contrast between the FLL and the parenchyma, this method is based on the hypothesis that the optimal reference frame for initialising the FLL occurs when highest contrast between different regions of the image is obtained – in particular between the regions of the FLL and the parenchyma. However, neither the FLL nor the parenchyma regions have yet been localised, or segmented, prior to detecting the reference frame, making the direct measurement of the contrast between them impossible. Therefore, all calculations are carried out across the entire area of the "workspace". This workspace is automatically segmented by employing brightness intensity change detection, as previously described (Bakas, et al. 2012). The proposed method then investigates how the variation of intensity across the workspace changes over time, after spatially discretising the workspace into local patches.

Specifically, the workspace is subdivided into J such patches, each of $q \times q$ pixels, where q is small compared with the overall size of the image, and the pixel brightness intensity values are averaged over each patch j, as follows;

$$I_j(t) = \frac{1}{R} \sum_{r=1}^{R} p_{j,r}(t) \tag{1}$$

where R is the area (in pixels) of each patch ($R = q^2$), and $p_{j,r}(t)$ is the intensity of the rth pixel included inside the jth patch in this frame t.

Identification of the frame of maximum contrast between the two ROIs is then carried-out by assessing the distribution of these local intensities, i.e. from the "spread" of the

average intensity values across all the patches. This spread can be measured using the standard deviation (*SD*). The *SD* of the patch intensity values within the workspace, for each frame *t*, is also known as the root mean square contrast (Peli 1990) and can be used as a useful quantitative measure of 'contrast'. Specifically, this metric does not depend on the spatial frequency content or the spatial distribution of contrast in the image, and is defined as:

$$SD(t) = \sqrt{\frac{1}{J}\sum_{j=1}^{J}\left(I_j(t) - I_{wsp}(t)\right)^2} \tag{2}$$

where $I_j(t)$ is the average brightness intensity of the jth patch in the current frame *t* and $I_{wsp}(t)$ is the global average intensity of the workspace at frame *t*, defined as:

$$I_{wsp}(t) = \frac{1}{L}\sum_{l=1}^{L} p_l(t) \tag{3}$$

where *L* is the total number of pixels within the workspace and $p_l(t)$ is the brightness intensity of the lth pixel in the current frame *t*.

We note short-term/high-temporal-frequency fluctuations in these values on a frame-by-frame basis, which are due to noise, and/or movements of the transducer, and/or patient respiratory motion (Fig.2(b)). These fluctuations are irrelevant to the dynamics within the workspace over a longer timescale. We are interested in low-temporal-frequency variations (over seconds, or many frames, rather than individual frames) in brightness, which depict the actual temporal changes in the brightness intensity (i.e. the dynamic behaviour of the ROIs) within the workspace. Hence, the average frame intensities ($I_{wsp}(t)$) are smoothed over a short timescale using a low-pass moving average filter (Averkiou, et al. 2003) centred on the frame of interest, given by the following formula:

$$\widehat{I_{wsp}}(t) = \frac{1}{2Q+1}\left[I_{wsp}(t+Q) + I_{wsp}\left(t+(Q-1)\right) + \cdots + I_{wsp}(t - Q)\right] \tag{4}$$

where, $I_{wsp}(t)$ is given by Eq.3, $\widehat{I_{wsp}}(t)$ is the corresponding smoothed value at frame *t*, and *2Q+1* is the length of the moving average filter (Fig.2(b)).

The frame of best contrast can then be located by studying how the quantity *SD(t)* changes over time and finding when it is maximal. This was carried out and evaluated for various values of *q* (and hence of *J*) as described in the Section "Defining Error Margins based on the Radiologist's Suggestion".

Ideally, the maximum contrast should be achieved when the maximum value of *SD(t)*, *SD_{max}*, occurs. In cases where *SD(t)* continues to increase throughout the CEUS cine-loop, this may be taken to be the *SD* in the final frame of the sequence. However, this choice of "best" reference frame, t_0, necessarily involves a compromise between obtaining the "best" possible contrast in the reference frame and having enough frames left in the sequence after t_0, in order to study the dynamic behaviour of the ROIs. Thus, for practical purposes, t_0 is chosen to be the time index of the first frame for which the standard deviation ($SD(t_0)$), reaches or exceeds a specified fraction (1-β) of *SD_{max}*, where β is a given number close to, but greater than, 0 (e.g. $\beta = 0.05$). Thus, this frame is chosen such that:

$$t_0 = \text{argmin}\{t: SD(t) \geq (1 - \beta)SD_{\max}\} \tag{5}$$

For β sufficiently small, the difference between the frames of *SD(t_0)* and *SD_{max}* should not be perceivable to the human eye.

Performance Evaluation Metric

To quantitatively evaluate the results of the proposed method for automatically selecting the optimal reference frame (t_0), we firstly define the contrast between the ROIs in a

frame t as the absolute difference between the average brightness intensity of the FLL (I_{FLL}) and that of the parenchyma (I_{Par}) at that frame as:

$$\Delta I(t) = |I_{FLL}(t) - I_{Par}(t)| \qquad (6)$$

We then use as an evaluation measure, the ratio (CR) of the contrast between the ROIs in a frame suggested by the proposed method (t), relative to the contrast between the same ROIs in the frame suggested by the radiologist (t_{RS}):

$$CR(t) = \frac{\Delta I(t)}{\Delta I(t_{RS})} \qquad (7)$$

where values of $CR(t)$ above 1 would indicate that the proposed method suggested a frame with higher contrast between the two ROIs than that in the frame specified by the radiologist, i.e. the FLL is more clearly visible in the automatically-selected frame. Conversely, values of $CR(t)$ lower than 1 would indicate that the proposed method performs worse than the radiologist, in terms of the contrast between the ROIs in the proposed reference frames.

This is considered a more meaningful evaluation measure than the difference (as an absolute number of frames) between t_{RS} and t_0 (Bakas, et al. 2013), since the rate of change of the enrichment seen in the image plane is dependent on the exact nature of an FLL and its dynamic behaviour, i.e. its enhancement properties, which affect the timescale for such discrepancies between the frame selected automatically and by the radiologist.

Evaluation of the Optimal Patch Size

To investigate the variation of brightness between different parts of the workspace, the latter was divided into rectangular local patches each of $q \times q$ pixels. Different values of q (i.e. $q \in \{1, 2, 3, \ldots, 32\}$) were used, to divide the workspace into local patches of various sizes, allowing study of the effect of moving from a fine-grained resolution (for $q=1$) to various coarser-grained resolutions (for $q>1$).

For evaluating the performance of the proposed method for each patch size $q \in \{1, 2, 3, \ldots, 32\}$, a 10-fold cross validation over 60 of the 64 clinical cases was performed using a model configuration comprising two sets (i.e. a training set and a test set) over the provided data. The training set of each fold f is used to estimate the optimal value of the parameter β for this fold (Eq.5). Values of the parameter β in the range [0, 0.2] were considered in each fold, increased using a step size of 0.01. Note that β was defined in Eq.5 and represents the largest fractional amount by which the SD of the pixel brightness intensities in the suggested reference frame may be less than the corresponding maximum (or final) value of SD (i.e. SD_{max}) such that the difference between these is considered to be not perceivable by the human eye. Thus, for any given fold and for each value of the parameter β, a reference frame is found by the proposed method, providing one value of the evaluation measure $CR(t)$ (Eq.7) for each case of the training set. Subsequently, for each value of the parameter β, the values of $CR(t)$ are averaged across cases. When all values of β in the specified range have been applied, the optimal value of β for each fold is selected as the one that maximises the average value of $CR(t)$ over the cases in the training set for this fold (see Table 1). After finding the value of parameter β that maximises $CR(t)$ for each fold by using only that training set, the corresponding test set is employed to evaluate this value of β and compute $CR(t)$ over the new (test) examples, not seen in the training set. This time, only that optimal value of parameter β for each fold was used with the corresponding test set, providing a $CR(t)$ score for each case of the test set. Similarly to before, these $CR(t)$ scores are averaged over the cases in that test set and the results are shown in the bottom row of Table 1.

Evaluating the Radiologist's Suggestion

Assuming that the radiologist attempts to identify the frame with the maximum contrast between the two ROIs, Eq.6 is used with the GT shape descriptors of the two ROIs over the duration of each sequence to identify the frame (t_{GT}), where the contrast between the

two ROIs reaches its actual global maximum. In current clinical practice, the radiologist's suggestion (t_{RS}) is considered to be the best decision for the reference frame. However, there is substantial uncertainty in the precise choice of this frame t_{RS}, even by a radiologist experienced in using CEUS, due to the decision being based on `visual inspection' of the CEUS cine-loop. To investigate this uncertainty and to identify any potential systematic error, after bearing in mind that the human eye accumulates information steadily over time (Zimmermann, et al. 2013), we evaluated the time difference between frame t_{RS} and t_{GT}, by modifying the evaluation measure of $CR(t)$ (Eq.7) as follows.

$$CR(t_{GT}) = \frac{\Delta I(t_{GT})}{\Delta I(t_{RS})} \tag{8}$$

where $\Delta I(t_{GT})$ is the maximum contrast between the two ROIs as obtained from the GT, in frame t_{GT}.

Table 1. Results of the 10-fold cross-validation used with the Spatiotemporal Intensity Variation approach. CR stands for 'contrast ratio', SD for 'standard deviation', and IQR for 'inter-quartile range'.

FOLD	1	2	3	4	5	6	7	8	9	10
Based on SD (Neighbourhood size q=27)										
Best β value	0.06									
CR_{Train}	1.266	1.261	1.194	1.276	1.256	1.295	1.263	1.237	1.092	1.256
CR_{Test}	1.004	1.05	1.647	0.912	1.089	0.744	1.025	1.261	2.571	1.093
Based on IQR (Neighbourhood size q=28)										
Best β value	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.12	0.11
CR_{Train}	1.193	1.202	1.132	1.22	1.199	1.239	1.196	1.165	1.029	1.216
CR_{Test}	1.053	0.969	1.602	0.809	0.994	0.64	1.024	1.309	2.387	0.846

Defining Error Margins based on the Radiologist's Suggestion

We need to define a "margin of error" ($ME(t)$), based on the radiologist's suggestion, as the largest acceptable difference (ε) from the value of 1 that it would be acceptable for the proposed method to return in the contrast ratio evaluation measure ($CR(t)$). Thus, we first estimate $ME(t_{GT})$ as:

$$ME(t_{GT}) = 1 - \frac{1}{CR(t_{GT})} \Rightarrow ME(t_{GT}) = 1 - \frac{\Delta I(t_{RS})}{\Delta I(t_{GT})} \tag{9}$$

The largest error (ε), defined in Eq.10 below, is needed to define a bound on the acceptable difference between $\Delta I(t)$ and $\Delta I(t_{RS})$.

$$|ME(t)| \leq \varepsilon \Rightarrow \frac{1}{1 + \varepsilon} \leq CR(t) \leq \frac{1}{1 - \varepsilon} \tag{10}$$

We are particularly interested in setting the lower bound of $CR(t)$, as frames chosen by the proposed method with higher contrast between the ROIs than that of the frames suggested by the radiologist should be considered better.

Results

Assessment of the Radiologist's Suggestion

The radiologist's suggestion is firstly evaluated by estimating the time difference between the reference frame t_{RS} and the frame with the actual maximum contrast between the

ROIs t_{GT}. The results show evidence for this systematic error, by finding t_{RS} occurring later in the sequence than t_{GT} in 60 out of 64 cases, with mean and median difference of 41.66 and 26 frames, respectively (Fig.3).
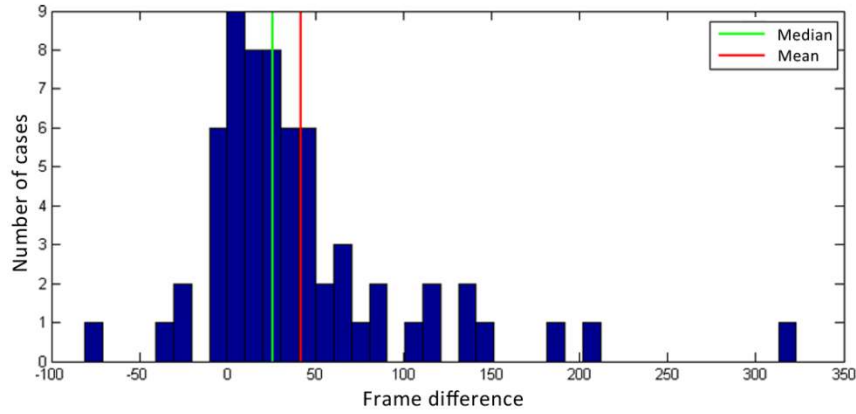

Figure 3. Distribution of the difference between the frames suggested by the radiologist (t_{RS}) and (t_{GT}). The fact that the frame difference is typically larger than 0 illustrates the radiologist's bias to select a later frame than the frame with the actual maximum contrast.

Furthermore, the radiologist's suggestion is compared against the actual maximum contrast between the ROIs, by applying Eq.8 to all 64 cases available are shown in Fig.4. Values of $CR(t_{GT})$ close to 1 indicate similar contrast between the frame suggested by the radiologist (t_{RS}) and the frame of the actual maximum contrast (t_{GT}). It is observed that three cases returned a value much higher than 1, indicating that a frame with much higher contrast between the two ROIs existed than that in the frame specified by the radiologist for those examples. Actual frames from these cases are shown in Fig.5.
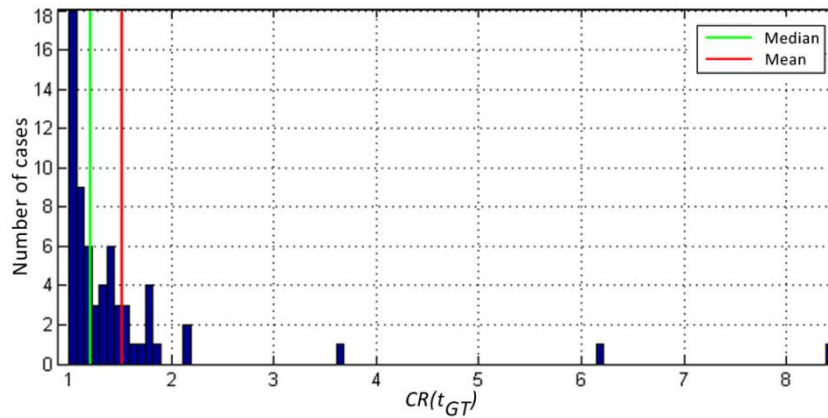

Figure 4. Distribution of the contrast ratio, in the reference frame suggested by the radiologist (t_{RS}) relative to that of the frame giving maximum contrast between the regions of interest (ROIs), as obtained from their ground truth (GT), i.e. at t_{GT}. Values of the contrast ratio (CR) between the ROIs at the time of its actual global maximum, $CR(t_{GT})$, close to 1 indicate similar contrast in frames t_{RS} and t_{GT}, whereas values larger than 1 indicates that a frame with higher contrast between the two ROIs exists than that in the frame specified by the radiologist.

Application of Eq.9 over all cases gives an average absolute value of ε equal to 0.23, which represents the average "error" (ε) of the radiologist's suggestion, currently considered to be the optimal decision (Fig.6). Consequently, by using this value of ε in Eq.10, we estimate this lower bound on $CR(t)$ as:

$$\frac{1}{1+\varepsilon} \le CR(t) \ \Rightarrow \ 0.813 \le CR(t) \tag{11}$$

Therefore, in the evaluation of the proposed method, any $CR(t)$ value above 0.813 should be considered acceptable.

(a) $t_{RS}=8$ (b) $t_{GT}=52$

(c) $t_{RS}=48$ (d) $t_{GT}=185$

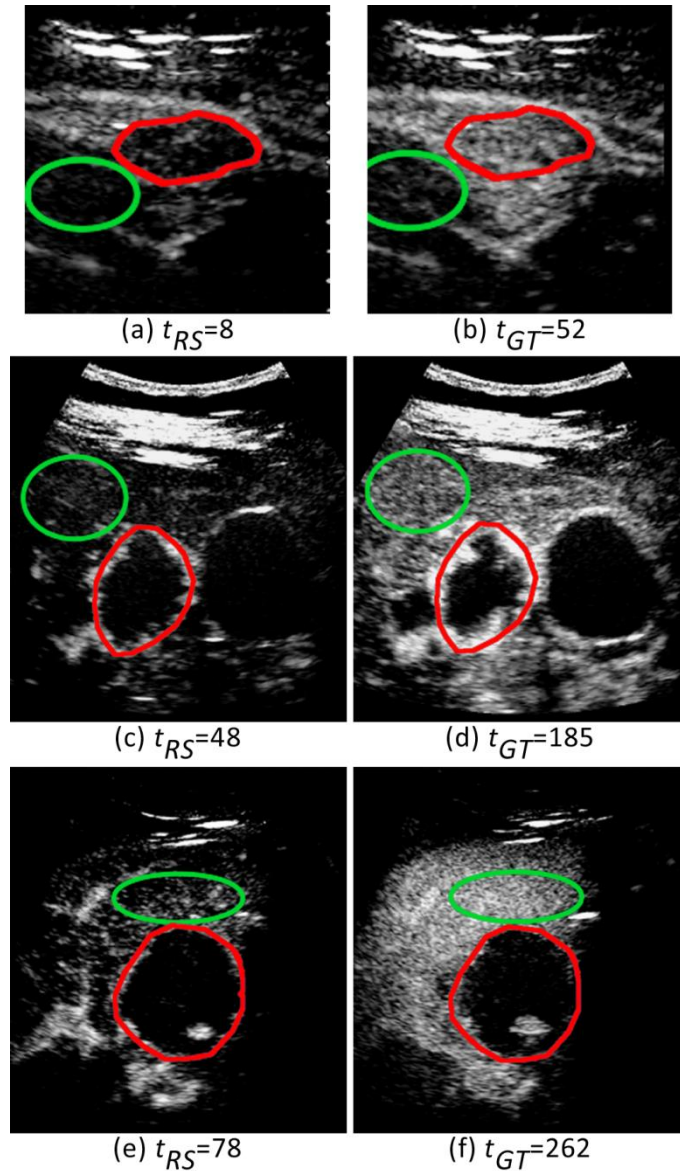(e) $t_{RS}=78$ (f) $t_{GT}=262$

Figure 5. Visual representation of the three cases (one in each row) that returned a value for the contrast ratio (CR) between the ROIs at the time of its actual global maximum, $CR(t_{GT})$ (Eq.8) much higher than 1, indicating that a frame exists with much higher contrast between the two regions of interest (t_{GT}, right column), than the one specified by the radiologist (t_{RS}, left column). The red and the green contours denote the focal liver lesion and the parenchyma regions, respectively.
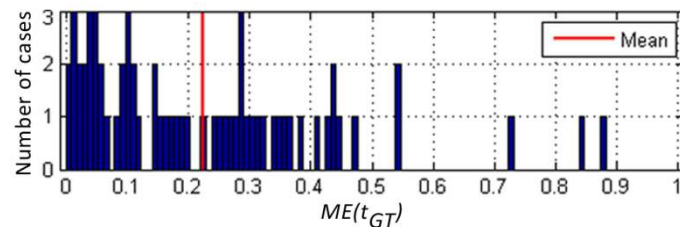

Figure 6. Distribution of the margin of error ($ME(t_{GT})$), defined by Eq.10, across all available data. The acceptable average error (ε) of the frames suggested by the radiologist (t_{RS}), and hence acceptable for the evaluation of the proposed methods, is equal to 0.23 and noted by the red vertical line.

Quantitative Evaluation of the Proposed Method

After applying all patch sizes q in the range {1, 2, 3, …, 32}, the results shown in Fig.7(a) indicate that the neighbourhood size $q=27$ gives the best performance, with average

$CR(t)$ score of 1.24 if the method is based on *SD*, such that 90% (i.e. 9 out of 10 folds = 54 out of 60 cases) of the test set were found to give $CR(t)$ values above the lower bound of acceptable $CR(t)$ scores as obtained from the assessment of the radiologist's suggestion, i.e. $CR(t) \geq 0.813$ (Eq.11). On the other hand, based on the Inter-Quartile Range (IQR), the patch size $q=28$ gives the best performance, with average $CR(t)$ score of 1.1636, with 80% (i.e. 8 out of 10 folds = 48 out of 60 cases) of the test set found to have CR scores above the lower bound of acceptable $CR(t)$ scores (see Table 1). Furthermore, the results presented in Fig.7(b) illustrate the performance of the proposed method for the optimal β value over all folds of the cross-validation and for each patch size q.
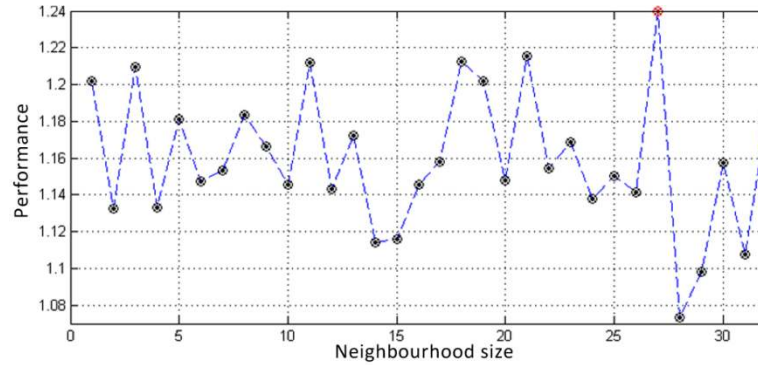
(a) Using the optimal β value for each fold
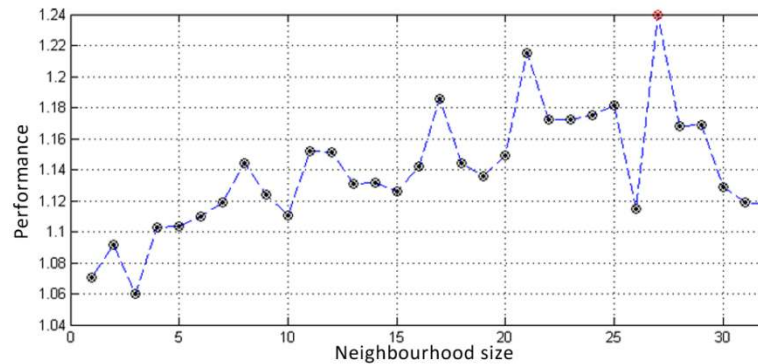
(b) β fixed at its overall optimal value

Figure 7. Performance of method (y-axis) for all different values of neighbourhood size (x-axis) attempted i.e. $q \in \{1, 2, 3, …, 32\}$. The performance is calculated by averaging contrast ration (CR) scores across the entire test set and across all 10 folds.

Clinical Value of the Proposed Method

Qualitative Assessment

To confirm the potential clinical value of the proposed method, which is based on assessing the spatiotemporal intensity variation in a CEUS workspace, all the automatically suggested reference frames were qualitatively evaluated by a radiologist who was not involved in the data acquisition. A binary decision (i.e. Accept/Reject) was provided, based on whether the FLL is well-distinguished and could be delineated given only this single reference frame. Furthermore, the reference frames suggested by i) the gradient-based method of assessing the whole workspace's intensity variation over time (Bakas, et al. 2013), ii) the radiologists involved in the data acquisition (t_{RS}), and iii) the frames with the actual maximum contrast between the FLL and the parenchyma (t_{GT}), were all also reviewed by the same independent radiologist (who was not involved in the data acquisition), in order to investigate the feasibility of identification and delineation of the FLL given only these single

frames and whether the criterion of maximum contrast is sufficient by itself for this task. The results of this review (see Table 2) suggest that a single frame, even if it is the reference frame provided by another radiologist (column RS), might not always be sufficient on its own to segment the FLL present, and it may be necessary to assess the dynamics included in a CEUS cine-loop to allow the required task to be carried out.

Fig.8 shows comparisons between suggested reference frames for initialising the region of an FLL, from the CEUS cine-loops of six example clinical cases.

Table 2. Qualitative evaluation of reference frames in the 64 provided clinical cases, by an independent radiologist. `RS' stands for `Radiologist's Suggestion', `GT' refers to the frames with the actual maximum contrast between the regions of interest, `SIV' stands for ' Spatiotemporal Intensity Variation', and `Gb' stands for the Gradient-based method of (Bakas, et al. 2013).

	RS	GT	SIV	Gb
Number of frames accepted	58/64	53/64	56/64	35/64
Percentage of cases accepted	90.63%	82.81%	87.5%	54.69%

Quantitative Assessment

To further evaluate the clinical significance of our method, we have also applied a computer-aided segmentation approach (Bakas, et al. 2015, Bakas, et al. 2014), to confirm the effect of the automatically selected frame on the segmentation of FLLs, and compare to the frames at t_{RS} and t_{GT}. The average segmentation accuracy i) on the frame selected by the radiologists (t_{RS}) was 95.2%, ii) on the automatically selected frame was 95.8%, and iii) on the frame with the actual maximum contrast (t_{GT}) was 96.1%. The distributions of the obtained results are shown in Fig.9. The obtained results confirm that our method comes in reasonable agreement with the radiologist, whereas in the current dataset the effect of the proposed method yielded a better segmentation of the lesion than this obtained using the frames at t_{RS}.
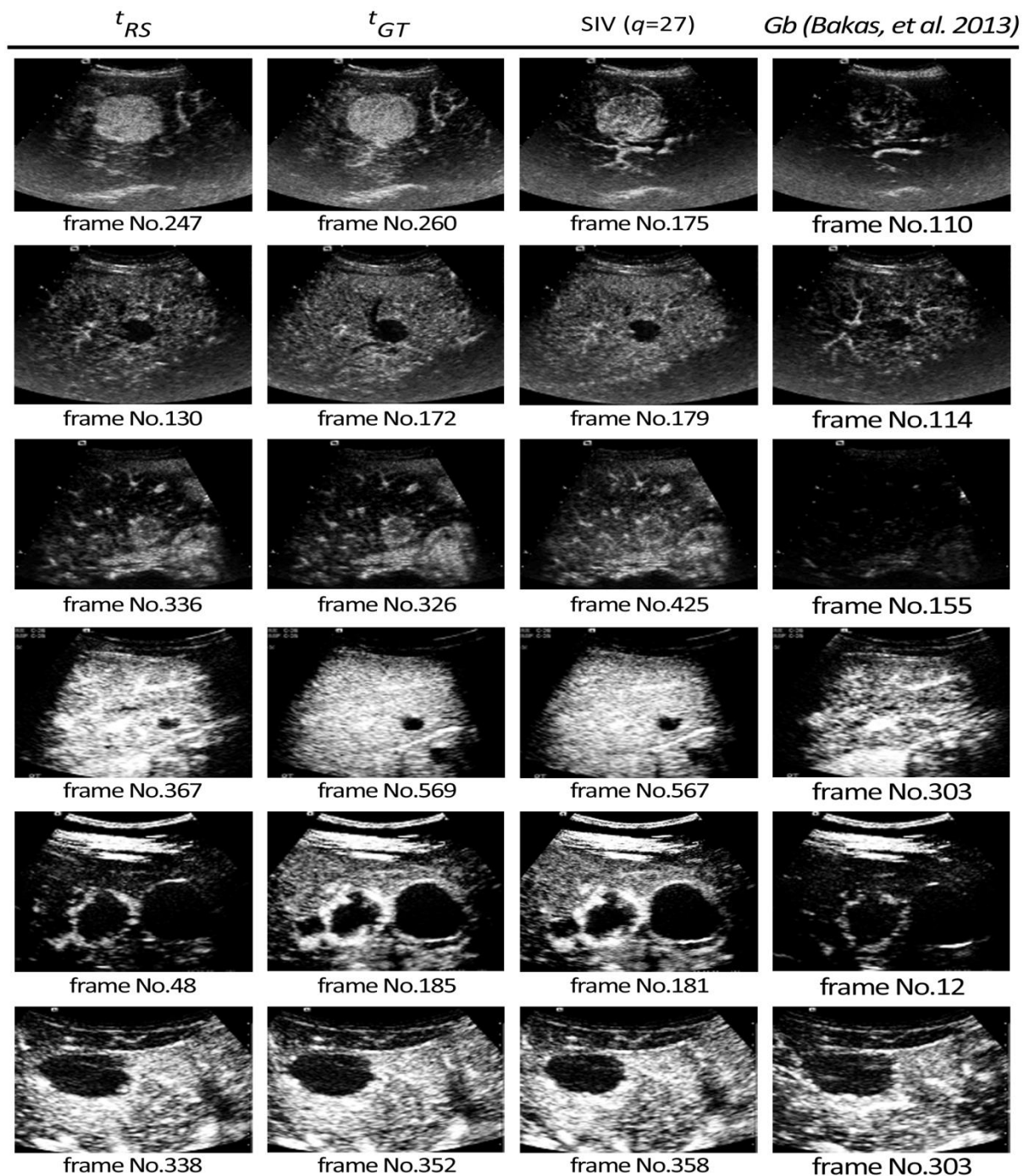
Figure 8. Visual representations of reference frames examples for the segmentation of the focal liver lesion (FLL) region, as suggested by an expert radiologist (t_{RS} – column 1), by the provided ground truth (t_{GT} – column 2), by the proposed method (SIV – column 3), and by the Gradient-based method (Gb – column 4) of (Bakas, et al. 2013). The figures shown are the actual video frames from the cine-loops of six of the provided clinical cases. The outcome of the Spatiotemporal Intensity Variation (SIV) method is based on the optimal neighbourhood size $q=27$ and $\beta=0.06$. It is noted that the frame suggested by the Gb is always earlier in the frame sequence than t_{RS}, whereas SIV provides a frame with even better contrast between the FLL and parenchyma regions than t_{RS} does.
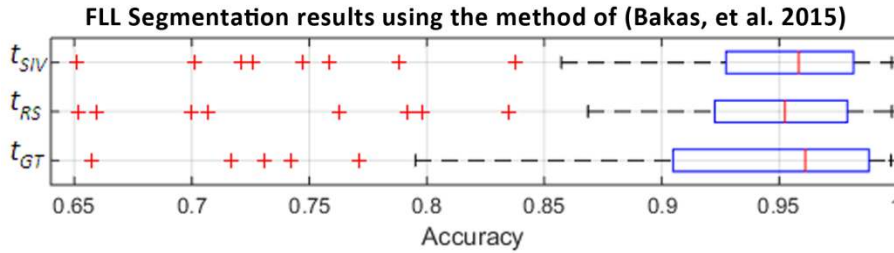
Figure 9. Comparison of the accuracy distributions on segmenting the apparent focal liver lesion (FLL) across all provided cases, for the frames selected automatically by the proposed method (t_{SIV}), the frames selected by the radiologist (t_{RS}), and the frames with the actual maximum contrast between the two ROIs (t_{GT}).

Discussion

We have proposed an automatic and deterministic method, based on statistical techniques and on obtaining high contrast between different regions of the image, to identify the optimal reference frame in a recording of a CEUS liver scan, enabling the delineation of an FLL, whilst avoiding error propagation due to suboptimal reference frame selection.

Considering that any human judgement, based on visual inspection, will necessarily involve some uncertainty, the proposed method, gives results in reasonable agreement with the choices of an expert radiologist by visual observation. This uncertainty was quantitatively shown when three example cases indicated that a frame with much higher contrast between the two ROIs existed, than that in the frame specified by the radiologist for those examples (Fig.5). It is observed that the first case, which depicts a malignant FLL, is of particular interest as the lower boundaries of the FLL are not clearly defined in frame $\boldsymbol{t_{RS}}$ (Fig.5(a)), while in frame $\boldsymbol{t_{GT}}$ (Fig.5(b)) it is the upper boundaries of the FLL that are not very clear. The other two cases each depict a haemangioma (benign FLL) and in both frames $\boldsymbol{t_{RS}}$ and $\boldsymbol{t_{GT}}$, the depicted FLL is sufficiently well-distinguished to segment it.

Quantitative evaluation on clinically-acquired multi-institutional CEUS liver scans was used to assess the level of confidence in the proposed method's suggestions by assessing the method's parameters. The very fine-grained resolution (i.e. small patch sizes q) might be affected by outliers occurring due to speckle noise, which are irrelevant to the ROIs, but can influence the distribution – and consequently the range and the SD – of the pixel brightness intensities over a given frame. These outliers are excessively bright or dark pixels that affect the distribution of intensities, exaggerating their complete range. Therefore, local spatial averaging using more coarse-grained resolutions was considered appropriate, averaging over groups of neighbouring individual pixels within a small patch, in order to avoid such effects due to noise whilst still taking into consideration the approximate location of individual pixels due to the location of the patch they are in. It should be mentioned that the SD as a measure of contrast is more robust to moderate levels of noise than other measures, such as the Michelson contrast, also known as *Visibility* (Michelson 1927), that depend on the full range of intensity values and are hence affected more by extreme values due to noise. Furthermore, the IQR is theoretically considered to be even more robust to effects due to outliers than is the SD. Specifically, outliers are not included in calculation of IQR, but are included when calculating the SD. IQR is the difference in brightness intensity values between the first and third quartiles, where these are defined to be the values in an ordered list of neighbourhood brightness intensities such that 25% of all the patches within the workspace are darker than the first quartile and another 25% are brighter than the third quartile. Therefore, in addition to studying how the SD of brightness changes over time, the changes in the IQR of brightness over time were also investigated. However, the practical results based on the IQR were shown to be inferior for our data to those obtained using the SD (Table 1).

The contrast ratio score obtained for the training set (CR_{Train}) having a large value does not necessarily mean that the optimal β parameter used for this fold is the optimal one overall, as the former might have been obtained through "overfitting" to the training data. The $CR(t)$ scores obtained for the training set are very likely to be higher than the more general $CR(t)$ score (the actual generalisation score) obtained when applying the method with the proposed β parameter value to new examples, not seen in the training set. This can be observed by comparing the CR_{Train} and CR_{Test} values for each fold in Table 1, where in most cases CR_{Test} is smaller than CR_{Train}. The overall performance of the method is then calculated by averaging the CR scores obtained for each test set across the different folds.

From the results shown in Fig.7(b), it can be observed that the proposed method generally performs better than the radiologist in selecting a suitable reference frame, irrespective of the patch size, as its average $CR(t)$ value is above 1 for every patch size tried in the range $1 \leq q \leq 32$. Furthermore, the larger the patch size, the less computational time is required for the method to be carried out on a given dataset. Although inspection of Fig.7(b) confirms the hypothesis that small patch sizes q normally underperform larger q values and a peak is shown at $q = 27$, the conclusion can be drawn that the performance of the method is relatively insensitive to changes in the precise value of the parameter q but the optimal values obtained might depend on the data provided.

Even though according to statements made by radiologists that the optimal reference frame is the one with the maximum contrast between the regions of the FLL and the parenchyma, the obtained results (column GT in Table 2) show that this is not always exactly true in a quantitative manner. Moreover, the results support that, when the frame with the maximum contrast is shown to a radiologist, this frame is not always considered as the most suitable. This is also the reason that the contrast ratio measure $CR(t)$, we used in our evaluation, was based on the radiologist's suggested frame t_{RS} instead of on the one with the actual maximum contrast t_{GT}, (see denominator in Eq.7). Finally, based on the average $CR(t)$ scores given in Section "Quantitative Evaluation of the Proposed Method" (Table 1), and using the number of accepted frames according to the binary decision of the radiologist (Table 2), it is clear that the proposed method (SIV) considerably outperforms the gradient-based method of (Bakas, et al. 2013) (Fig.10), and the automatically identified reference frames of SIV are comparable to those suggested by a radiologist for the purpose of delineating the FLL and the parenchyma (Table 2).
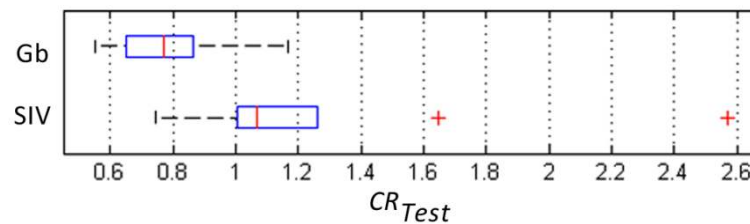

Figure 10. Comparison of performance between the Gradient-based method (Gb) of (Bakas, et al. 2013) and the currently proposed method, Spatiotemporal Intensity Variation (SIV), for automatically identifying the optimal reference frame, using the CR_{Test} value.

Finally, the clinical value of the proposed method was assessed both qualitatively and quantitatively. The qualitative review of the frames suggested by the proposed method (t_{SIV}), the radiologist (t_{RS}) and the frames with the actual maximum contrast between the two ROIs according to the GT (t_{GT}), revealed that complete segmentation of an FLL using just a single frame may not always be feasible, and further review of the cine-loop should be considered in some cases. However, the results of this review (Table 2) show that our method performs very similarly to an expert radiologist. Furthermore, application of a computer-aided segmentation method in all these aforementioned frames (t_{SIV}, t_{RS}, t_{GT}) has shown that the proposed method provides frames that can be in reasonable agreement to the ones suggested

by the current gold standard (i.e. the expert radiologist) but also that the frames suggested at t_{SIV} can be considered optimal for the segmentation of FLLs (Fig.9).

Conclusions

Offline assessment of CEUS cine-loops is essential for diagnosis, staging, treatment planning, and follow-up of FLLs. However, training to perform CEUS examinations and interpretation of their data is time-consuming and subjective. Automation of such tasks in a reliable manner would improve the current clinical practice for the assessment and evaluation of FLLs. Towards this end, we have proposed an automatic and deterministic method, based on statistical techniques, to identify the optimal reference frame in a recording of a CEUS liver scan, enabling the delineation of an FLL, whilst avoiding error propagation due to suboptimal reference frame selection. Experimental results are in reasonable agreement with the choices of expert radiologists, as shown by both quantitative and qualitative evaluation.

The clinical value of this study falls into the creation of a standardised criterion for producing repeatable and reproducible results for automatically temporally localising an FLL. Therefore, the radiologists' confidence will be increased, for rapidly, more easily and objectively identifying the frame where the contrast between the regions of the FLL and the parenchyma is maximal. Hence in this frame the FLL is best-distinguished from the parenchyma, and well-defined in the image plane, allowing for the segmentation of its boundaries (Bakas, et al. 2015, Bakas, et al. 2014) and the subsequent quantification of its vascular enhancement pattern (Anaye, et al. 2011, Bakas, et al. 2012, Bakas, et al. 2014, Christofides, et al. 2016, Dietrich, et al. 2012, Quaia 2011, Rognin, et al. 2006, Rognin, et al. 2010). Furthermore, this technique could be used as a training tool for junior radiologists, promoting the wider use of CEUS by non-specialist centres, potentially leading to reduced turnaround times, lower cost to healthcare services and less distress to patients and their families (Lanka, et al. 2007, Westwood, et al. 2013).

Acknowledgments

References

Albrecht T, Blomley M, Bolondi L, Claudon M, Correas J-M, Cosgrove D, Greiner L, Jager K, Jong Nd, Leen E, Lencioni R, Lindsell D, Martegani A, Solbiati L, Thorelius L, Tranquart F, Weskott HP, Whittingham T. Guidelines for the Use of Contrast Agents in Ultrasound - January 2004. Ultraschall in der Medizin 2004; 25:249-56.

Anaye A, Perrenoud G, Rognin N, Arditi M, Mercier L, Frinking P, Ruffieux C, Peetrons P, Meuli R, Meuwly JY. Differentiation of Focal Liver Lesions: Usefulness of Parametric Imaging with Contrast-Enhanced US. Radiology 2011; 261:300-10.

Averkiou M, Powers J, Skyba D, Bruce M, Jensen S. Ultrasound Contrast Imaging Research. Ultrasound Q 2003; 19:27-37.

Bakas S, Chatzimichail K, Hunter G, Labbe B, Sidhu PS, Makris D. Fast Semi-Automatic Segmentation of Focal Liver Lesions in Contrast-Enhanced Ultrasound, Based on a Probabilistic Model. TCIV Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 2015; ePub-ahead-of-print.

Bakas S, Hoppe A, Chatzimichail K, Galariotis V, Hunter G, Makris D. Focal Liver Lesion Tracking in CEUS for Characterisation based on Dynamic Behaviour. Springer, Advances in Visual Computing, Lecture Notes in Computer Science (LNCS) 2012; 7431:32-41.

Bakas S, Hunter G, Thiebaud C, Makris D. 2013 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. *IEEE 9th Int. Conf. on Intelligent Environments (IE)*. Athens, Greece, 196-203.

Bakas S, Labbe B, Hunter GJA, Sidhu PS, Chatzimichail K, Makris D. 2014 Fast Segmentation of Focal Liver Lesions in Contrast-Enhanced Ultrasound Data. *Medical Image Understanding and Analysis*. City University, London, UK, , 73-78.

Bakas S, Sidhu PS, Sellars ME, Hunter GJA, Makris D, Chatzimichail K. 2014 Non-invasive Offline Characterisation of Contrast-Enhanced Ultrasound Evaluations of Focal Liver Lesions: Dynamic Assessment Using a New Tracking Method. *European Congress of Radiology (ECR)*.

Brannigan M, Burns PN, Wilson SR. Blood Flow Patterns in Focal Liver Lesions at Microbubble-Enhanced US. Radiographics 2004; 24:921-35.

Chiorean L, Cantisani V, Jenssen C, Sidhu P, Baum U, Dietrich C. Focal masses in a non-cirrhotic liver: The additional benefit of CEUS over baseline imaging. European journal of radiology 2015; 84:1636-43.

Christofides D, Leen E, Averkiou MA. Evaluation of the Accuracy of Liver Lesion DCEUS Quantification With Respiratory Gating. IEEE Transactions on Medical Imaging 2016; 35:622-29.

Claudon M, Dietrich CF, Choi BI, Cosgrove DO, Kudo M, Nolsoe CP, Piscaglia F, Wilson SR, Barr RG, Chammas MC, Chaubal NG, Chen M-H, Clevert DA, Correas JM, Ding H, Forsberg F, Fowlkes JB, Gibson RN, Goldberg BB, Lassau N, Leen ELS, Mattrey RF, Moriyasu F, Solbiati L, Weskott H-P, Xu H-X. Guidelines and Good Clinical Practice Recommendations for Contrast Enhanced Ultrasound (CEUS) in the Liver - Update 2012: A WFUMB (World Federation for Ultrasound in Medicine and Biology) - EFSUMB (European Federation of Societies for Ultrasound in Medicine and Biology) Initiative in Cooperation with Representatives of AFSUMB (Asian Federation of Societies for Ultrasound in Medicine and Biology), AIUM (American Institute of Ultrasound in Medicine), ASUM (Australasian Society for Ultrasound in Medicine), FLAUS (Latin-American Federation of Societies for Ultrasound in Medicine and Biology) and ICUS (International Contrast Ultrasound Society). Ultrasound in Medicine & Biology 2013; 39:187–210.

Dietrich C, Averkiou M, Correas J-M, Lassau N, Leen E, Piscaglia F. An EFSUMB introduction into Dynamic Contrast-Enhanced Ultrasound (DCE-US) for

quantification of tumour perfusion. Ultraschall in der Medizin-European Journal of Ultrasound 2012; 33:344-51.

Dietrich CF, Kratzer W, Strobel D, Danse E, Fessl R, Bunk A, Vossas U, Hauenstein K, Koch W, Blank W, Oudkerk M, Hahn D, Greis C. Assessment of Metastatic Liver Disease in Patients with Primary Extrahepatic Tumors by Contrast-Enhanced Sonography Versus CT and MRI. World Journal of Gastroenterology 2006; 12:1699–705.

Ding H, Wang WP, Huang BJ, Wei RX, He NA, Qi Q, Li CL. Imaging of Focal Liver Lesions: Low-Mechanical-Index Real-time Ultrasonography With SonoVue. J Ultrasound Med 2005; 24:285-97.

Harvey CJ, Blomley MJK, Eckersley RJ, Cosgrove DO. Developments in Ultrasound Contrast Media. Eur Radiol 2001; 11:675–89.

Lanka B, Jang HJ, Kim TK, Burns PN, Wilson SR. Impact of Contrast-Enhanced Ultrasonography in a Tertiary Clinical Practice. J Ultrasound Med 2007; 26:1703–14.

Llovet JM. Updated Treatment Approach to Hepatocellular Carcinoma. Journal of Gastroenterology 2005; 40:25–35.

Llovet JM, Burroughs A, Bruix J. Hepatocellular Carcinoma. Lancet 2003; 362:1907–17.

Michelson A. Studies in Optics: Chicago, Ill. : The University of Chicago Press, 1927.

Mischi M, Rognin N, Averkiou M. Ultrasound imaging modalities. 2014.

Nicolau C, Vilana R, Catala V, Bianchi L, Gilabert R, Garcia A, Bru C. Importance of Evaluating All Vascular Phases on Contrast-Enhanced Sonography in the Differentiation of Benign from Malignant Focal Liver Lesions. American Journal of Roentgenology 2006; 186:158-67.

Peli E. Contrast in Complex Images. J Opt Soc Am A 1990; 7:2032-40.

Quaia E. Microbubble ultrasound contrast agents. European Radiology 2007; 17:1995-2008.

Quaia E. Assessment of tissue perfusion by contrast-enhanced ultrasound. European Radiology 2011; 21:604-15.

Quaia E. The real capabilities of contrast-enhanced ultrasound in the characterization of solid focal liver lesions. European Radiology 2011; 21:457-62.

Quaia E. Solid focal liver lesions indeterminate by contrast-enhanced CT or MR imaging. Abdominal imaging 2012; 37:580-90.

Rognin N, Campos R, Thiran J, Messager T, Broillet A, Frinking P, Mercier L, Arditi M. 2006 A new approach for automatic motion compensation for improved estimation of perfusion quantification parameters in ultrasound imaging. *Eighth French Conference on Acoustics (Tours, France)*, 61-5.

Rognin NG, Arditi M, Mercier L, Frinking PJ, Schneider M, Perrenoud G, Anaye A, Meuwly J-Y, Tranquart F. Parametric imaging for characterizing focal liver lesions in contrast-enhanced ultrasound. IEEE transactions on ultrasonics, ferroelectrics, and frequency control 2010; 57:2503-11.

Schneider M. Charasteristics of SonoVue. Echocardiography 1999; 16:743-46.

Seitz K, Strobel D. A Milestone: Approval of CEUS for Diagnostic Liver Imaging in Adults and Children in the USA. Ultraschall Med 2016; 37:229-32.

Sidhu PS. 2014 Are We Using Enough CEUS in Clinical Practice? Role In Radiation Dose Reduction. *European Congress in Radiology (ECR)*. Vienna, Austria.

Sirli R, Sporea I, Martie A, Popescu A, Danila M. Contrast Enhanced Ultrasound in Focal Liver Lesions - a Cost Efficiency Study. Medical Ultrasonography 2010; 12:280–85.

Skyba DM, Price RJ, Linka AZ, Skalak TC, Kaul S. Direct In Vivo Visualization of Intravascular Destruction of Microbubbles by Ultrasound And Its Local Effects On Tissue. Circulation 1998; 98:290-93.

Sporea I, Sirli R, Martie A, Popescu A, Danila M. How useful is contrast enhanced ultrasonography for the characterization of focal liver lesions. J Gastrointestin Liver Dis 2010; 19:393-98.

Strobel D, Seitz K, Blank W, Schuler A, Dietrich CF, Herbay Av, Friedrich-Rust M, Bernatik T. 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 2009; 30:376–82.

Sugimoto K, Shiraishi J, Moriyasu F, Doi K. Computer-Aided Diagnosis of Focal Liver Lesions by Use of Physicians' Subjective Classification of Echogenic Patterns in Baseline and Contrast-Enhanced Ultrasonography. Academic Radiology 2009; 16:401-11.

Westwood ME, Joore MA, Grutters JPC, Redekop WK, Armstrong N, Lee K, Gloy V, Raatz H, Misso K, Severens J, Kleijnen J. Contrast-Enhanced Ultrasound Using SonoVue R (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 2013; 17:1-243.

Wilson SR, Burns PN. An Algorithm for the Diagnosis of Focal Liver Masses Using Microbubble Contrast-Enhanced Pulse-Inversion Sonography. AJR Am J Roentgenol 2006; 186:1401-12.

Wilson SR, Burns PN. Microbubble-Enhanced US in Body Imaging: What Role? Radiology 2010; 257:24–39.

Zimmermann E, Morrone MC, Burr DC. Spatial Position Information Accumulates Steadily Over Time. J Neurosci 2013; 33:18396-401.