

©2019, Elsevier. Licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International <http://creativecommons.org/about/downloads>



Evaluation of Indirect Methods for Motion Compensation in 2D Focal Liver Lesion Contrast-Enhanced Ultrasound (CEUS) Imaging

Authors:

Spyridon Bakas^{a,b},

Matthaios Doulgerakis-Kontoudis^{a,c},

Gordon J.A. Hunter^a,

Paul S. Sidhu^d,

Dimitrios Makris^a,

Katerina Chatzimichail^e

^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 *Medical Imaging and Image Interpretation Group, School of Computer Science, University of Birmingham, Edgbaston, B15 2TT, UK*

^d *Department of Radiology, King's College Hospital, Denmark Hill, London, SE5 9RS, United Kingdom*

^e *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@uphs.upenn.edu

Abstract

This study investigates the application and evaluation of existing *indirect methods*, namely point-based registration techniques, for the estimation and compensation of observed motion included in the 2D image plane of Contrast-Enhanced Ultrasound (CEUS) cine-loops recorded for the characterization and diagnosis of focal liver lesions (FLL). The value of applying motion compensation in the challenging modality of CEUS is to assist the quantification of the perfusion dynamics of an FLL in relation to its parenchyma, allowing for a potentially accurate diagnostic suggestion. Towards this end, this study also proposes a novel quantitative multi-level framework for evaluating the quantification of FLLs, which to the best of our knowledge remains undefined, notwithstanding many relevant studies. Our results suggest the ‘compact and real-time descriptor’ as the optimal indirect motion compensation method in CEUS, following a quantitative evaluation of nineteen other indirect algorithms and configurations, while also considering the requirement for computational efficiency.

Keywords:

Contrast-enhanced Ultrasound,

Focal Liver Lesions,

Computer-Aided Detection,

Motion Compensation,

Automation,

Liver Cancer,

Ultrasound,

Contrast Agents,

CEUS

Introduction

Contrast-Enhanced Ultrasound (CEUS) is an important imaging technique used for assessing focal liver lesions (FLLs), with the ability to differentiate benign and malignant tumors. CEUS is an extension of the ultrasound (US) modality, which has inherent high temporal resolution, but the CEUS examination requires an intravenous (IV) injection of a contrast agent (Harvey, et al. 2001, Schneider, et al. 1995) that accentuates the returning echoes (increased brightness) from blood and with the ability to differentiate patterns of enhancement of FLL compared to surrounding healthy liver (i.e., the parenchyma) in the image frames. Interpretation of CEUS cine-loops is time-consuming, labor-intensive and subjective, as it is based on manual offline/post-examination assessment by expert radiologists following the CEUS guidelines by the World Federation for Ultrasound in Medicine and Biology (WFUMB) (Claudon, et al. 2013).

Considering the beneficial impact that CEUS can have in FLL evaluation and the bottleneck imposed by the cumbersome interpretation of its recorded cine-loops, computer-aided automation of the process may be advantageous. Various attempts have been made towards such automation, including the identification of the optimal reference frame in the acquired CEUS cine-loop (Bakas, et al. 2013, Bakas, et al. 2017b), the delineation of the CEUS image plane (i.e., the sector-shaped area including the ultrasonographic image) (Bakas, et al. 2012b), as well as the delineation of the FLL (Bakas, et al. 2017a). Once these “*initialization*” tasks have been performed, monitoring the relative changes of brightness intensity of the different regions of interest (ROIs) over time (i.e., the perfusion temporal dynamics) is of particular importance, as such information leads to the vascular signature (VS) of individual FLLs that forms the basis of diagnostic decisions.

Monitoring these perfusion dynamics through a 2D CEUS cine-loop entails many different challenges, primarily due to the nature of the US examination (Tang, et al. 2011) which includes much unpredictable motion, and depends on the experience of the operator, as well as the patient’s underlying physiological status. The IV injection of contrast-enhancing agents results in dramatic changes of tissue appearance during the examination (Fig.1), when captured in a non-linear contrast-specific mode that cancels

linear tissue echoes. Appearance variations may also depend on noise, intensity changes, or out-of-plane movements of the captured structures that lead to changes of their apparent location, orientation and/or size. Such movements may be due to the free-hand acquisition approach (i.e., the relative motion between a hand-held transducer and the liver), as well the patient's inner organ deformation and physiological motion due to cardiac movements and breathing (Derek, et al. 2001). For this reason, the standard care protocol suggests a fixed sagittal view to maintain breathing motion in-plane during an examination, refraining from changing any acquisition parameters, such as the zoom level or the field of view. In addition, to minimize the effect of such motion, the patients are asked to breath shallowly, which can result in a breathing pattern with low amplitude but still rather irregular movement, such that no motion pattern that could be directly exploited for motion compensation is present (Wollny, et al. 2012). The continuous irregular repetitive nature of all these disturbances inevitably degrades the quality of the acquired data, and if no action is taken to compensate for the observed motion, the estimated VS of the FLL(s) may be inaccurate, as it may consider intensity from pixels corresponding to different tissues, and hence may lead to inaccurate diagnostic suggestions. Therefore, the main challenge and primary aim for automatically assessing the FLL perfusion dynamics is to compensate for the apparent motion and without any acquisition standards (e.g., patient's breath-hold) being instructed by the radiologists beforehand, beyond the ones used on standard clinical practice (Claudon, et al. 2013).



Figure 1. Example frames from an FLL CEUS cine-loop illustrating the dramatic appearance changes of the liver and the FLL.

Towards this end, this study focuses on compensating for the observed FLL motion within the 2D CEUS image plane and through a CEUS cine-loop using automated motion estimation methods. Motion

estimation approaches are divided into two major categories (Makela, et al. 2002); the *direct* (Irani and Anandan 2000) and the *indirect* methods (Torr and Zisserman 1999). The former recover information directly from measurable image quantities at each pixel in the image plane, based on either the intensity itself or its gradient (i.e., the edge-based energy), for estimating motion vectors in successive frames of a sequence. These algorithms are also known as correlation or template matching algorithms (Althof, et al. 1997). Examples of direct methods are based on block matching (Rognin, et al. 2006, Ta, et al. 2012), phase correlation (Castro and Morandi 1987), and normalized cross-correlation or its modifications (Ding, et al. 2001). Direct methods share the ‘intensity constancy constraint’, where the intensity of an initialized ROI must be constant over a sequence of frames, and are sensitive to noise and illumination changes (Irani and Anandan 2000). Therefore, considering the local and global dramatic intensity changes over the duration of a CEUS recording due to the use of contrast-enhancing agents, such methods may not be considered optimal for estimating motion in CEUS cine-loops. On the other hand, the latter, *indirect* methods employ a sparse set of distinct geometric image features, matching across successive frames, to estimate motion over a sequence. These geometric image features describe local structural information and are either manually or automatically selected “landmarks” (e.g., corners (Harris and Stephens 1988) or lines (Rothwell, et al. 1995)) robustly characterized for saliency by statistical multi-dimensional local feature descriptors. Specifically, point-based registration techniques (PBRTs) are considered as the most suitable methods for automatic detection of “landmarks” when illumination changes occur, and in multi-modal analysis (e.g., MRI/US (Correas 2014)), mainly due to their invariance to changes in illumination, scale, orientation, contrast and viewpoint (Khajone, et al. 2014, Torr and Zisserman 1999).

The aim of this study is to facilitate the automated accurate quantification of the FLL’s perfusion in CEUS scans and hence allow for an accurate diagnostic suggestion. In previous work, we have used the Scale Invariant Feature Transform (SIFT) as a PBRT alone (Bakas, et al. 2012b), as well as coupled with a direct method (Bakas, et al. 2011, Bakas, et al. 2012a), to compensate for motion in CEUS cine-loops. The contributions of the present study are two-fold; a) an extensive investigation of a wide range of existing PBRTs (Ambai and Yoshida 2011, Bay, et al. 2006, Dalal and Triggs 2005, Evans 2009, Harris and Stephens

1988, Lowe 2004, Rosten and Drummond 2005, Rosten and Drummond 2006, Rublee, et al. 2011, Shi and Tomasi 1994, Vedaldi and Fulkerson 2010), and their configurations, and the identification of the optimal one for motion compensation in FLL CEUS cine-loops, and b) a novel quantitative multi-level framework of metrics for evaluating motion compensation techniques in CEUS scans involving FLLs. Results are computed and compared for all applied PBRTs, based on the proposed evaluation protocol, demonstrating the value of the applied methods as a useful advisory tool for a radiologist to forge his own evaluation.

Background

This section reviews existing work relevant to the scope of the present study. Firstly, the standard clinical workflow for FLL characterization in CEUS is presented, followed by existing computational studies, considering approaches without and with motion compensation. The latter are all based on direct methods, which we argue that they might not be appropriate for CEUS cine-loops. For this reason, we review indirect methods that have not been applied for motion compensation in FLL CEUS cine-loops by other researchers before. The matching strategies and distance metrics used by these methods are also presented.

Standard Clinical Workflow for FLL Characterization

A CEUS examination starts with the identification of a FLL, and then proceeds onto a targeted CEUS assessment of the lesion over several minutes, initially (60-90 seconds) recording the examination on a cine-loop, with static images thereafter. Following this acquisition, a radiologist assesses the acquired CEUS cine-loop “offline” by an exhaustive visual inspection, assessing the vascular pattern and sequence of lesion enhancement in the arterial phase and throughout the recorded cine-loop, and any static images from the late/portal venous phase of the examination, eventually characterizing the FLL (Burrowes, et al. 2017, Claudon, et al. 2013, Dill-Macky, et al. 2002, Wilson and Burns 2010, Wilson, et al. 2018). Besides being time consuming, this workflow does not produce a quantitative analysis, requires considerable effort and the expert knowledge of experienced and specially-trained radiologists, yet still prone to misinterpretation as diagnostic results are dependent on experience for interpretation of the data. Computational studies for assessing the quantification of FLLs using CEUS are discussed, below.

Computational Studies Without Motion Compensation

Approaches of this type perform analysis of CEUS cine-loops after i) excluding frames with motion, or ii) using only static images, or iii) acquiring the data following strict instructions being given to the radiologists beforehand.

Lemke et al. (2004) conducted a prospective study, where patients previously scanned by CT and MRI were re-scanned with US and CEUS, aiming to clinically evaluate the diagnostic accuracy of CEUS in comparison with CT and MRI. Static images were obtained for each patient at predefined intervals and, after having the FLL(s) manually annotated, the discrete signal of each FLL's VS was used to assist with classification.

Huang-Wei et al. (2006) analyzed cine-loops of the arterial phase, with an approximate duration of 7-12 seconds, of CEUS examinations using the QontraXt software (AMID, Italy). Information was obtained after performing background subtraction (Piccardi 2004), and the use of a parametric image to improve the visualization. However, background subtraction is expected to work satisfactory only if the transducer is static and the patient's breathing is negligible. Also, the background (i.e., the parenchyma) intensity values change continuously, further constraining the applicability of background subtraction. Considering that brightness intensity reflects the log compression used for creating the CEUS image, and that the QontraXt software does not linearize pixel amplitude, rendering background subtraction is potentially a problematic/hazardous operation. Furthermore, only a particular type of FLL (i.e., Focal Nodular Hyperplasia) was included in their analysis and without motion correction.

Shiraishi et al. (2008) developed a computer-aided diagnostic scheme for the classification of three very specific types of FLLs with hyper-enhanced VS (i.e., Metastasis, Hemangioma, and Hepatocellular Carcinoma). FLL delineation was obtained manually in each frame of the acquired cine-loops by a physician, obviously labor-intensive and time consuming. A cascade of six independent Artificial Neural Networks was then used for the task of classification.

Goertz et al. (2010) performed CEUS quantification of FLLs after excluding 77 case studies out of 110 due to various motion-related challenges, e.g., ‘wiggly’ recording, patient’s fragile holding of breath, FLL getting lost in depth. The Sonoliver software (V.2.6.8/2.7.0 – Tomtec, Unterschleissheim, Germany) was used, for semi-automatic motion compensation by combining manual delineation of the FLL and the parenchyma, background subtraction, and automatic alignment of ROIs. The outcome of their study implies that Sonoliver could only work with data obtained according to particular acquisition standards, where motion is minimal, due to the use of background subtraction.

Computational Studies with Motion Compensation Using Direct Methods

Rognin et al. (2006) described an approach to automatically compensate for rigid motion in CEUS cine-loops and obtain perfusion curves by iteratively maximizing a pixel-based similarity criterion, using steepest gradient descent optimization. This approach requires substantial manual input for: i) selecting a reference frame, ii) delineating an analysis region within the CEUS image, iii) delineating the FLL, and iv) delineating the parenchyma. The similarity criterion used was mutual information (MI) (Shekhar and Zagrodsky 2002), based on intensities of the manually delineated image plane. Although MI has been shown to be useful (accurate and robust) when attempting to register frames from different modalities (Lu, et al. 2008, Maes, et al. 1997, Pluim, et al. 2001), it cannot be considered optimal when the intensity and the apparent shape of the ROIs change dramatically over time, as the ‘intensity constancy constraint’ criterion is not met and such pixel-based analysis is susceptible to noise (Irani and Anandan 2000). Another study (Anaye, et al. 2011) has also used multiple region annotation but for parametric maps of perfusion curves, and to differentiate between various FLLs.

Ta et al. (2012) have shown that a pixelwise (i.e., fine-grained resolution) analysis, for CEUS cine-loops, is more informative for FLL diagnostics than ROI-based analysis. However, their main purpose was to compare differences between the dynamics of benign and malignant FLLs, as well as to quantify the spatial heterogeneity of FLLs, by performing a pixelwise analysis within manually initialized ROIs. Motion compensation was performed by maximization of MI of the manually initialized ROI between successive frames, in a similar manner to the approach of (Rognin, et al. 2006).

Such approaches based on “*direct*” methods give results that are neither repeatable nor reproducible, as they are heavily dependent on manual input.

Indirect Motion Estimation Methods

Indirect (or feature-based) methods register a sparse set of either manually or automatically selected “landmarks”, e.g., points, or lines, as distinct geometric image features between two frames, in order to estimate the apparent motion. Since no lines are detectable in our datasets, we focus entirely on PBRT methods.

Once the “landmark” points are detected, a statistical multi-dimensional local feature descriptor is employed to robustly characterize a “landmark” as a salient point at a higher level, to facilitate matching the corresponding landmark point in another frame, leading to the estimation of a motion vector between them. These single “point/vector” representatives of distinct geometric image features are generally described in the literature as “interest points”. Ideally, such points should be distinct, spread over the whole image, easily distinguishable amongst other points within the image plane, detectable in both frames and sufficiently stable to avoid influence by slight unexpected feature variation due to changes in the imaging conditions, i.e., noise, intensity variations, scaling, or rotation.

A variety of PBRTs have been introduced for this purpose (Ambai and Yoshida 2011, Bay, et al. 2006, Dalal and Triggs 2005, Harris and Stephens 1988, Leutenegger, et al. 2011, Lowe 2004, Mikolajczyk and Schmid 2005, Shi and Tomasi 1994) and most of them employ statistical multi-dimensional descriptors (e.g., in 64, 128, or 256 dimensions) to characterize the saliency of each “interest point” robustly and find more efficient correspondences in the descriptor space than in the image space, while being rendered independent of intensity constancy constraints. PBRTs are usually applied when local structural information is more meaningful than brightness intensity. PBRTs, compared to direct methods, do not require an initialized ROI, and hence can allow ‘*global*’ transformation estimation, providing information for what is usually called ‘*image registration*’ (Sotiras, et al. 2013). Such a statistically robust feature can itself be treated as a landmark with subpixel accuracy. Additionally, salient points found in one image without any

correspondences in the other image should not affect the performance of the registration (Fitzpatrick and West 2001).

Existing Point-Based Registration Techniques (PBRTs)

In 1988, the Harris corner detector (Harris and Stephens 1988) was introduced as one of the earliest mathematically rigorous point extraction techniques, where the extraction of a point is based on eigenvalues (λ_1, λ_2). The Harris detector has an extremely high repeatability score (Schmid, et al. 2000), defined as the number of corresponding points detected in both images, over the total number of observed points overall. Note that the repeatability criterion conflicts with that of localization, as smoothing might improve the former but deteriorate the latter (Canny 1986).

In 1994, Shi & Tomasi (1994) suggested another technique where the extraction of a point, similarly to Harris' method, is based on eigenvalues. The two methods differ in the "interest point" selection criterion, but share the same two limitations: i) lack of a multi-dimensional descriptor; ii) extraction of points at the pixel level instead of a subpixel resolution. According to (Schmid, et al. 2000), if Shi & Tomasi's method is combined with a multi-dimensional descriptor (e.g., the 'Histogram of Oriented Gradients', HOG), then the result is expected to give higher saliency points than Harris' method.

In 2004, the Scale Invariant Feature Transform (SIFT) (Lowe 2004) was proposed, which uses a set of Difference of Gaussians (DoG) filters at different scales as a method to extract "interest points" invariant to scale, orientation, illumination changes and camera viewpoint. Furthermore, SIFT provides a 128-dimensional descriptor, inspired by a model of biological vision, i.e., a complex of neurons in the primary visual cortex. These neurons react to gradients in specific orientations and spatial frequencies, where the gradient can change on a relatively small field. The outcome is a set of highly distinctive descriptors providing robust matches and being invariant to scale, rotation, noise, illumination, and viewpoint.

In 2005, the HOG (Dalal and Triggs 2005) was proposed as a descriptor extraction method that works in a similar manner to SIFT, but using a denser grid of overlapping blocks over a detection window.

Normalization of gamma, contrast and color is also performed, and the HOGs are extracted for each block and combined in a local descriptor over a detection window.

In 2006, the Features from Accelerated Segment Test (FAST) approach (Rosten and Drummond 2005, Rosten and Drummond 2006) was introduced, as a high-speed test for identifying salient points, with the intention to use it in real-time applications. Its potential speed efficiency is due to the use of decision trees to form a set of rules for deciding if a potential “interest point” should be kept.

In 2006, the Speeded Up Robust Features (SURF) (Bay, et al. 2006) method was presented as a scale- and rotation-invariant point detector and descriptor that works similarly to SIFT. However, SURF has lower computational cost than SIFT, especially for its descriptor extraction, due to the use of the ‘integral image’ (Viola and Jones 2004) for convolution operations, the extraction of points using a measure based on the Hessian matrix, and the use of a shorter (64 dimensions) descriptor.

In 2011, the Binary Robust Invariant Scalable Keypoints (BRISK) technique (Leutenegger, et al. 2011) was proposed, as a scale-invariant ROI and point detector. Detection was based on the FAST optimization, and localization was achieved by quadratic function fitting. A binary descriptor is formed by concatenating results of intensity comparisons, selected in such a way as to maximize saliency. Scale and rotation normalization make the BRISK descriptors invariant to distortions of scale and orientation.

Also in 2011, the Oriented FAST and Rotated Brief (ORB) approach (Rublee, et al. 2011) was published, based on FAST and BRIEF (Calonder, et al. 2010). This feature algorithm uses FAST (or Harris) in pyramids to detect salient points based on the “cornerness” response. The orientation of these points is then assessed by first-order moments and BRIEF descriptors are obtained, rotated according to the measured orientation.

The Compact And Real-time Descriptors (CARD) method (Ambai and Yoshida 2011) was also proposed in 2011 as another point detector and descriptor that also works similarly to SIFT. Its extension over SIFT is that initially a multi-scale image pyramid is employed and then “interest points” are obtained

using a corner detector (Shi and Tomasi 1994). Although CARD avoids using DoG filters for time efficiency, it manages to efficiently detect salient scale-invariant points. The CARD descriptors are extracted using a combination of SIFT and Gradient Location-Orientation Histogram (GLOH) (Mikolajczyk and Schmid 2005) methodologies. Two speed improvements are proposed in CARD: i) bilinear interpolation of pixel values is avoided; ii) look-up table use for the computation of oriented gradients. The descriptors are converted to binary using a binary hashing function, leading to highly salient descriptors extracted in very low computational time.

Matching Strategies

Typically, a strategy is needed to match the statistical descriptors of salient points between two frames. After considering an evaluation of matching strategies for statistical descriptors (Mikolajczyk and Schmid 2005) comprising i) a threshold-based approach, ii) a nearest-neighbor based approach and iii) the nearest-neighbor distance ratio (NNDR), this study utilizes only the latter for all experiments. In the first approach, two points $(p_i(t_1), p_j(t_2))$ at frames t_1 and t_2 are accepted as a *matched correspondence* if the distance between their descriptors $(D(p_i, t_1)$ and $D(p_j, t_2))$ is below a certain threshold, where $D(p_i, t)=[w_{1,i}(t), w_{2,i}(t), \dots, w_{N,i}(t)]$ and $D(p_j, t)=[w_{1,j}(t), w_{2,j}(t), \dots, w_{N,j}(t)]$ are both N dimensional vectors, respectively, describing the two points p_i and p_j at time t . Thus, a point's descriptor may have several matches in another frame and some of these will be incorrect correspondences. In the second approach, points are accepted as corresponding if their descriptors are nearest-neighbors and the distance between them is below a specified threshold. In this second case a point's descriptor should have fewer false matches and hence higher precision. The third approach (NNDR) extends the second by applying the threshold to the distance ratio between the two first nearest-neighbors, and hence keeps only robust correspondences.

$$\frac{\|D(p_i, t_1) - D(p_j, t_2)\|}{\|D(p_i, t_1) - D(p_k, t_2)\|} < T \quad (1)$$

where T is the threshold value and $D(p_j, t_2)$, $D(p_k, t_2)$ are respectively the first and second nearest-neighbors to $D(p_i, t_1)$.

Distance Metrics

All above matching strategies are based on the application of a threshold to a measurable “distance” between two different descriptors and, depending on their nature, different metrics can be used to measure the distance between descriptors. The neighborhood of each point is used through the descriptors to rank potential matches by estimating the Euclidean distance between two descriptors as:

$$d_{D(p_i,t_1),D(p_j,t_2)} = \sqrt{\sum_{m=1}^N (w_{m,i}(t_1) - w_{m,j}(t_2))^2} \quad (2)$$

Alternatively, by considering the descriptors ($D(p_i,t_1)$ and $D(p_j,t_2)$) as points in the descriptor space with position vectors of unit length, then the Euclidean distance between them is closely related to their inner (or “dot”) product, which in turn provides a measure of the angle between these vectors. For small angles ($\leq 30^\circ$), it can be shown that the Euclidean distance between two such unit position vectors is analogous to the angle between them. Therefore, the dot product of the two descriptors may be used as an alternative to the Euclidean distance, as calculating dot products between unit vectors can be more computationally efficient than finding Euclidean distances.

$$d_{\vec{D}(p_i,t_1),\vec{D}(p_j,t_2)} \sim (\vec{D}(p_i,t_1) \cdot \vec{D}(p_j,t_2)) \quad (3)$$

Finally, the Hamming distance (Norouzi, et al. 2012) can be used to measure the ‘distance’ between two binary strings of the same length as the number of positions where the corresponding strings are different (i.e., the number of substitutions required to change from one string to the other). This offers a high computational efficiency in finding the distance between binary descriptors.

Materials

Equipment

This study utilizes multi-institutional data from the Department of Diagnostic Radiology of King’s College Hospital at Denmark Hill, London, UK and from the Radiology & Imaging Research Centre of Evgenidion Hospital, part of the National and Kapodistrian University of Athens in Greece. Both datasets were acquired using ACUSON Sequoia US Systems (Siemens, Mountain View, CA) with curvilinear

transducers (UK: S2000, 4~6MHz, GR: C512, 6-2MHz). All the GR cases were captured at the same frame rate of 25 frames per second (fps), whereas the UK cases were captured at different frame rates (namely 6-25 fps). The examinations in both datasets performed employing a low (≤ 0.1) mechanical index (AIUM. 2000, Ding, et al. 2005) technique (Cadence Contrast Pulse Sequencing, Siemens, Mountain View, CA), in order to avoid breaking the contrast-enhancing agents (Claudon, et al. 2013, Skyba, et al. 1998). The mechanical index roughly indicates and is related to the acoustic pressure of the US system (Averkiou, et al. 2003, Tiemann, et al. 1999).

All examinations included administration of the second-generation contrast medium, called SonoVue (Schneider 1999) (Bracco S.p.A., Milan, Italy), in a 2.4 ml bolus intravenous injection (at maximum 1 ml per second into an arm vein). SonoVue consists of sulphur hexafluoride microbubbles stabilized by a phospholipid monolayer shell. This contrast medium has demonstrated excellent depiction of FLL vascularity and perfusion (Wilson and Burns 2010), and generally improves the detection and characterization of liver and renal masses (Correas 2014).

Additional specific acquisition settings (e.g., selected scan plane, field depth, acoustic gain) of the equipment for each patient are unknown, as they were set by each radiologist individually at the start of each examination. These settings of the US equipment, as well as the scanning mode, have an important role in avoiding artefacts (Dietrich, et al. 2011), which could jeopardize the success of an examination. The acquisition method of the provided data reflects true clinical practice (Claudon, et al. 2013) and leads to increased variability within the datasets, as well as to a wide variety of different lengths of the cine-loops.

Additionally, all the algorithms tested in this study run in Matlab v.7.12 without parallelism, on an Intel i7-2620M platform with 8GB RAM.

Data

The retrospective multi-institutional CEUS cine-loops used in this study describe 64 FLL cases ($n_{UK}=50$, $n_{GR}=14$). Local Ethics Board approval waivers are attained in King's College Hospital (UK) and Evgenidion Hospital (GR), where maintenance of randomization codes for each case and hence

confidentiality of patients' records was secured by anonymizing the screened data. Consent for using the recordings and diagnostic results for research purposes was acquired from all patients.

The median patient age was 50 years (range: 19-91 years). Each clinical case includes a CEUS digital cine-loop with median duration of 24.35 seconds (range_{UK}: 3.1-89.83 seconds, range_{GR} 5.96-19.64 seconds) of the arterial and the portal-venous phases of the CEUS examinations, with at least one static image of the late phase. The image size for the UK Dataset is 1024×768 pixels, while for the GR Dataset is 768×576 pixels. Examples of 44 benign and 20 malignant FLLs were included in both datasets.

Each examination/acquisition within the UK Dataset was performed by one of two experienced radiologists with respectively nineteen and fourteen years of experience in use of CEUS, whilst all cases in the GR Dataset were obtained by a single radiologist with fifteen years of experience using CEUS. The radiologists acquired all data following the standard clinical practice (Claudon, et al. 2013), and without being instructed any additional specific criteria nor knowing that these data will be subsequently processed by computational tools. As well as being carried out retrospectively, analysis of the data was also performed “blind” to clinical diagnoses. Each cine-loop of the provided data included only a single FLL within the liver of each patient, of diameter between 0.5 and 8 cm. Also, the physical condition of the patients (e.g., proportion of body fat) varied considerably, which therefore increases the intrinsic variability of the provided data.

Ground Truth

The ground truth (GT) labels were manually delineated by a radiologist, at every frame t , for the FLL (i.e., $F_{GT}(t)$) and the liver (i.e., $L_{GT}(t)$) areas, based on the local intensity and shape information varying over the duration of the cine-loop. Then, the GT of the parenchyma (i.e., $P_{GT}(t)$) was obtained, which is defined as the relative complement of $F_{GT}(t)$ in $L_{GT}(t)$, i.e., the set of pixels in $L_{GT}(t)$, but not in $F_{GT}(t)$ (Fig.2).

The GT provided for the evaluation of this study was distinguished into two different types, i.e., rigid and non-rigid. The former was provided for all 64 cases and describe rigid transformations of the initialized

contours in frame t_0 across subsequent frames. The latter were provided for 11 cases of the UK Dataset only and includes non-rigid deformations of the shapes of the ROIs throughout the duration of each cine-loop.

In addition to the ROI shape descriptors, the diagnostic decisions for all cases of both datasets were provided retrospectively. These ‘diagnostic decisions’ describe the clinical classification of each FLL as either benign or malignant. In our study, the “Reference Standard” (RS) is used to validate the applied methods at the level of the clinical decision, i.e., decision level. RS defines a term used when additional imaging (e.g., MRI or CT) and consensus between radiologists were used as the means for diagnosing an FLL.

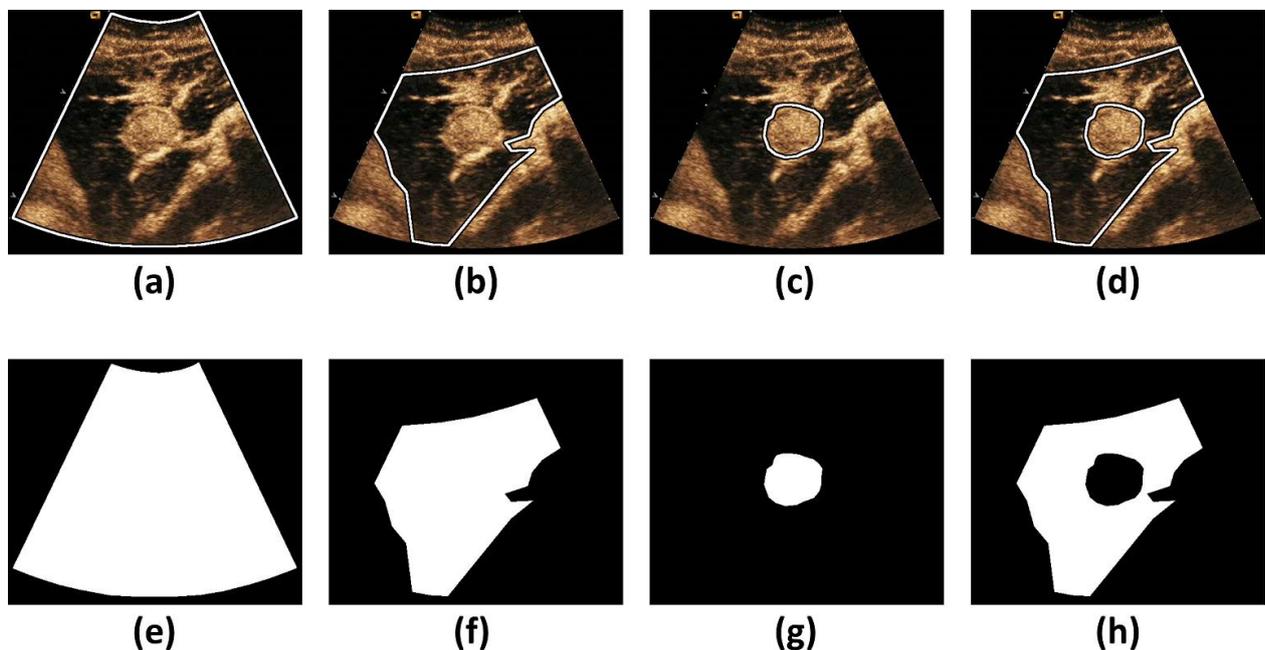


Figure 2. Manually annotated contour of different ROIs in an example reference frame (a-d), and the corresponding binary masks of their silhouettes (e-h). Left to right we note the CEUS image plane, the liver ($L_{GT}(t)$), the FLL ($F_{GT}(t)$) and the parenchyma ($P_{GT}(t)$).

Proposed Methods

This section describes the technical details of our proposed contributions. Specifically, in subsection “Tracking FLL Contours in CEUS Cine-loops” we describe a framework to use PBRTs for motion

estimation and compensation in FLL CEUS cine-loops. Then, our quantitative evaluation methodology is described, where we introduce a novel quantitative framework to evaluate the performance of motion compensation in CEUS at three different levels: pixel, behavioral, and decision. Based on this framework, we evaluate the complexity of transformations essential to capture the apparent motion in a CEUS cine-loop, as well as the computational efficiency of the applied PBRTs. Finally, we identify an appropriate compromise between the required frame-rate and the performance of a computational method for real-time motion compensation.

We applied a range of various existing PBRT configurations and implementations (S.Table 1) in 2D CEUS cine-loops and compared them according to the proposed evaluation criteria. Specifically, for ‘Harris corner detector’ (Harris and Stephens 1988) and the method of (Shi and Tomasi 1994), the MATLAB built-in implementations were used. Note that the original methods of ‘Harris’ and ‘Shi & Tomasi’ are not equipped with a statistical multi-dimensional local feature descriptor, therefore we individually combined them with the HOG descriptor (Dalal and Triggs 2005) as implemented in OpenCV and subpixel accuracy is achieved by applying an existing routine (Bradski 2000) before and after application of the descriptor. The PBRT of SIFT (Lowe 2004), using an 128-dimensional descriptor, was applied using the officially provided implementation by the author, as well as by OpenCV (Bradski 2000) and VLfeat (Vedaldi and Fulkerson 2010). The FAST (Rosten and Drummond 2005, Rosten and Drummond 2006), with and without subpixel accuracy, and the ORB (Rublee, et al. 2011), using a 32-dimensional descriptor, were used through OpenCV implementations. For SURF (Bay, et al. 2006), both its OpenCV and OpenSURF (Evans 2009) implementations were used, while considering both 64- and 128- dimensional descriptors. Last, but not least, CARD (Ambai and Yoshida 2011) was used through the official implementation, provided by the authors, while considering 64-, 128- and 256-bits descriptor.

Tracking FLL Contours in CEUS Cine-loops

PBRTs are tested in this study, to capture the changes in the FLL and the parenchyma contours over the frames of the provided CEUS cine-loops based on sparse optical flow and local feature descriptors (Fig.3).

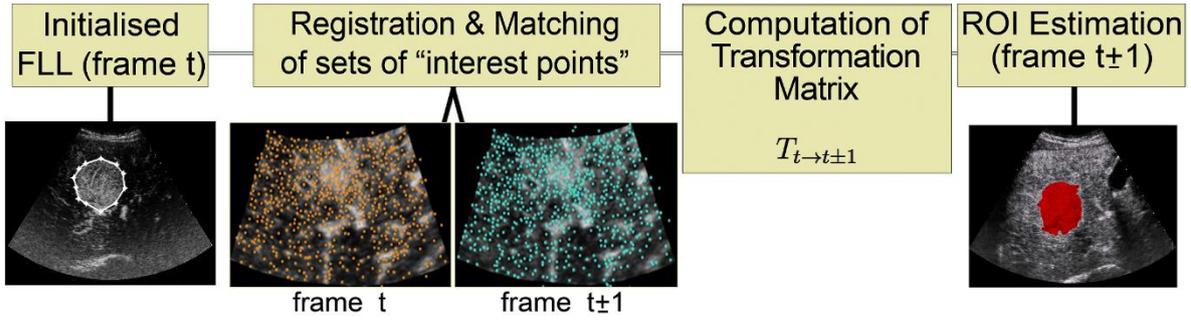


Figure 3. Pipeline of the motion estimation approach based on indirect methods. Where $t \in \{1, \dots, N\}$, $N =$ number of frames.

The FLL contour at frame t is described by $\Psi_{\text{FLL}(t)}$, which is a $2 \times K$ matrix, comprising K number of 2D points as the points of the FLL contour. The transformations of these contours for two consecutive frames t and $t+1$ can be estimated by localizing and registering matching sets of salient points $Q(t)$ and $Q(t \pm 1)$. $Q(t)$ is a $2 \times V(t)$ matrix, comprising $V(t)$ 2D salient points $q_v(t)$ at frame t , where $v \in [1, V(t)]$. Equivalently, $Q(t+1)$ is a $2 \times V(t+1)$ matrix, comprising $V(t+1)$ 2D salient points $q_\gamma(t+1)$ at frame $t+1$, where $\gamma \in [1, V(t+1)]$. In order to register matching sets of points, each localized 2D point $q_v(t)$ and $q_\gamma(t \pm 1)$ is assigned a descriptor of N -dimensions, $D(q_v, t)$ and $D(q_\gamma, t \pm 1)$, respectively.

The localized correspondences between the two frames are then estimated by minimizing the NNDR between $D(q_v, t)$ and $D(q_\gamma, t \pm 1)$ and selecting the nearest point in the descriptor space. Different metrics may be used for finding the distance between two descriptors, in the descriptor space, depending on the PBRT applied. Specifically, for every point $q_v(t)$, another point $q_\lambda(t \pm 1)$, where $\lambda \in [1, V(t \pm 1)]$, is found that their descriptors fulfil the following equation:

$$q_\lambda(t \pm 1) = \underset{q_\lambda(t \pm 1) \in Q(t \pm 1)}{\operatorname{argmin}} [\arccos(D^T(q_v, t) \cdot D(q_\lambda, t \pm 1))] \quad (4)$$

Similarly, a second point $q_{\lambda'}(t \pm 1)$ is found, where $\lambda' \in [1, V(t \pm 1)]$ and $\lambda' \neq \lambda$, such that:

$$q_{\lambda'}(t \pm 1) = \underset{q_{\lambda'}(t \pm 1) \in Q(t \pm 1)}{\operatorname{argmin}} [\arccos(D^T(q_v, t) \cdot D(q_{\lambda'}, t \pm 1))] \quad (5)$$

It is then accepted that there is a correspondence between points $q_v(t)$ and $q_\lambda(t \pm 1)$ in the two frames t and $t \pm 1$, if:

$$\frac{\arccos(D^T(q_v, t) \cdot D^T(q_\lambda, t \pm 1))}{\arccos(D^T(q_v, t) \cdot D^T(q_{\lambda'}, t \pm 1))} < Z \quad (6)$$

where $Z \in [0, 1]$ is a threshold value (Lowe 2004, Mikolajczyk and Schmid 2005). If a point $q_\lambda(t \pm 1)$ is the best match for more than one point $q_v(t)$, then all correspondences of $q_\lambda(t \pm 1)$ are rejected. Consequently, only reliable correspondences are registered.

These registered correspondences can assist in approximating the 2D geometric transformations on the image plane, i.e., Rotation, Translation, Euclidean, Similarity, Affine, Projective. The transformation matrix of these approximations $T_{t \rightarrow t \pm 1}$ is computed by the following equation:

$$U(t \pm 1) = T_{t \rightarrow t \pm 1} U(t) \Rightarrow T_{t \rightarrow t \pm 1} = U(t \pm 1) U(t)^{-1} \quad (7)$$

where $U(t)$ and $U(t \pm 1)$ are $2 \times R(t, t \pm 1)$ matrices, comprising $R(t, t \pm 1)$ 2D registered correspondences between frame t and $t \pm 1$, respectively. As matrix $U(t)$ is not square, $U(t)^{-1}$ is its Moore-Penrose inverse (Penrose 2008), computed using the singular value decomposition.

$T_{t \rightarrow t \pm 1}$ is then applied to $\Psi_{FLL(t)}$, in order to obtain their corresponding contours in the subsequent frame, estimated as:

$$\Psi_{FLL}(t \pm 1) = T_{t \rightarrow t \pm 1} \Psi_{FLL}(t) \quad (8)$$

Therefore, the contours are tracked in subsequent frames, while their shape and size can also be updated based on the transformation of subsequent frames.

Evaluation Methodology

Novel Multi-Level Quantification Framework

Quantitative evaluation at three levels, i.e., pixel-, behavioral- and decision-level, is performed, to assess both the technical contribution and the clinical value of the tested methods. The first two levels of evaluation are used to assess the effectiveness of the computer-aided motion compensation aspect, whilst the decision level is used to assess the effectiveness of the ability of the proposed methods for computer-aided diagnosis.

All PBRTs are firstly evaluated at the pixel level by calculating the accuracy of the proposed methods' decision of segmented ROIs in terms of the mean overlap metrics of i) Jaccard coefficient (J_t), ii) Precision (P_t), and iii) Recall (R_t), across the frames of the provided sequence. Secondly, the segmented ROIs

are validated at the behavioral level, to investigate the level of confidence of the methods' decision, as their signals are the means that the radiologists use to make a diagnosis. For this purpose, we introduce, Signal Quantification Metrics (SQM) that are based on the difference between the signal obtained from the manually annotated GT ($Signal_G$) and the signal obtained from the automated decision ($Signal_d$). The FLL VS signal is obtained by subtracting the perfusion curve of the parenchyma from this of the FLL.

The first SQM proposed here is the Maximum Signal Difference Ratio (MSDR) and is described by the ratio of the maximum normalized difference between $Signal_G$ and $Signal_d$, to $Signal_G$ in a frame of the cine-loop (Fig.4).

$$MSDR = \max_t \left| \frac{Signal_G(t) - Signal_d(t)}{Signal_G(t)} \right| \quad (9)$$

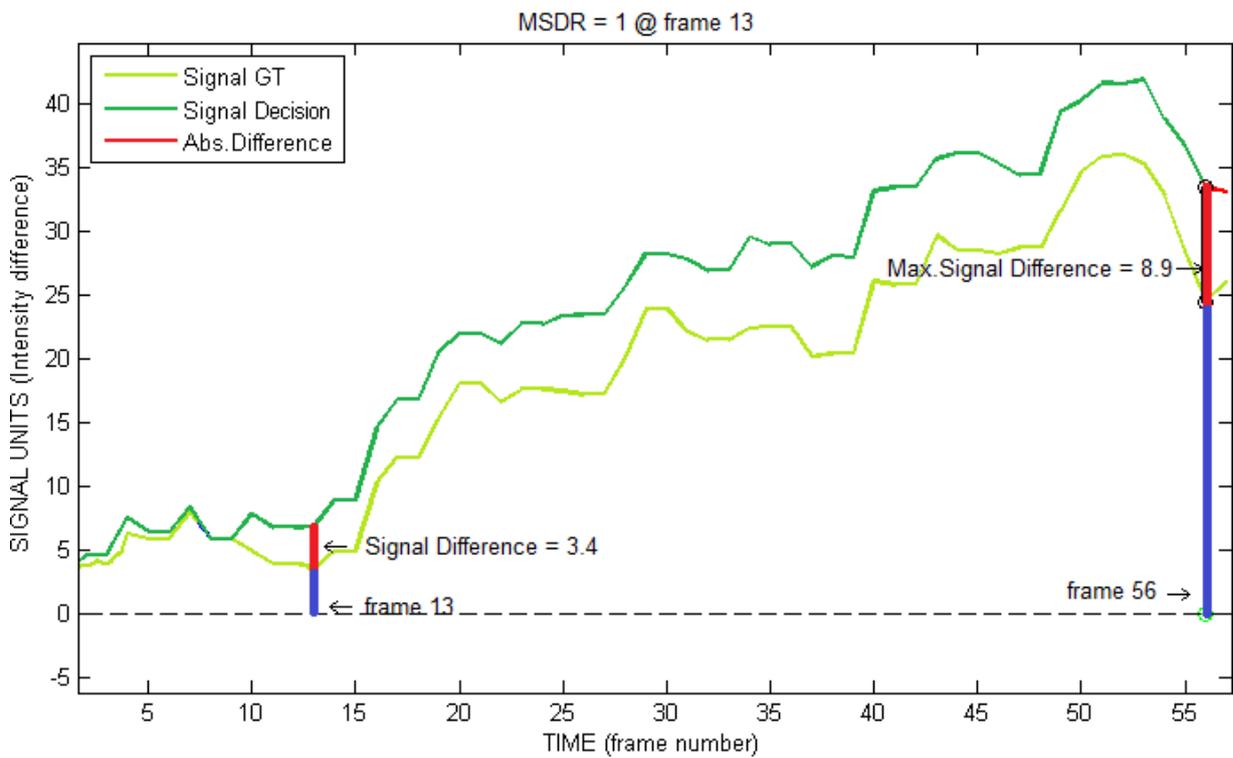


Figure 4. Example of the Maximum Signal Difference Ratio (MSDR). The red lines indicate the absolute signal difference, between $Signal_G$ (light green) and $Signal_d$ (dark green) for the FLL VS, and the blue line indicates the value of $Signal_G$ at these same points in time. Note that although the maximum signal difference occurs at frame 56 with 8.9 units of difference, the MSDR occurs at frame 13 (MSDR=1), even though the signal difference at this frame is 3.4 units.

The second SQM considered is the Normalized Cumulative Signal Difference (NCSD), which describes the ratio of the difference between the integrals of $Signal_G$ and $Signal_d$, over the time period of interest, to the integral of $Signal_G$ (Fig.5).

$$NCSD = \frac{\int_{t_i}^{t_f} (|Signal_G - Signal_d|) dt}{\int_{t_i}^{t_f} Signal_G dt} \tag{10}$$

where t_i and t_f define the time of the initial and the final frames, respectively. It is worth noting that in contrast with the pixel level overlap metrics smaller values for each of these SQM metrics correspond to better performance.

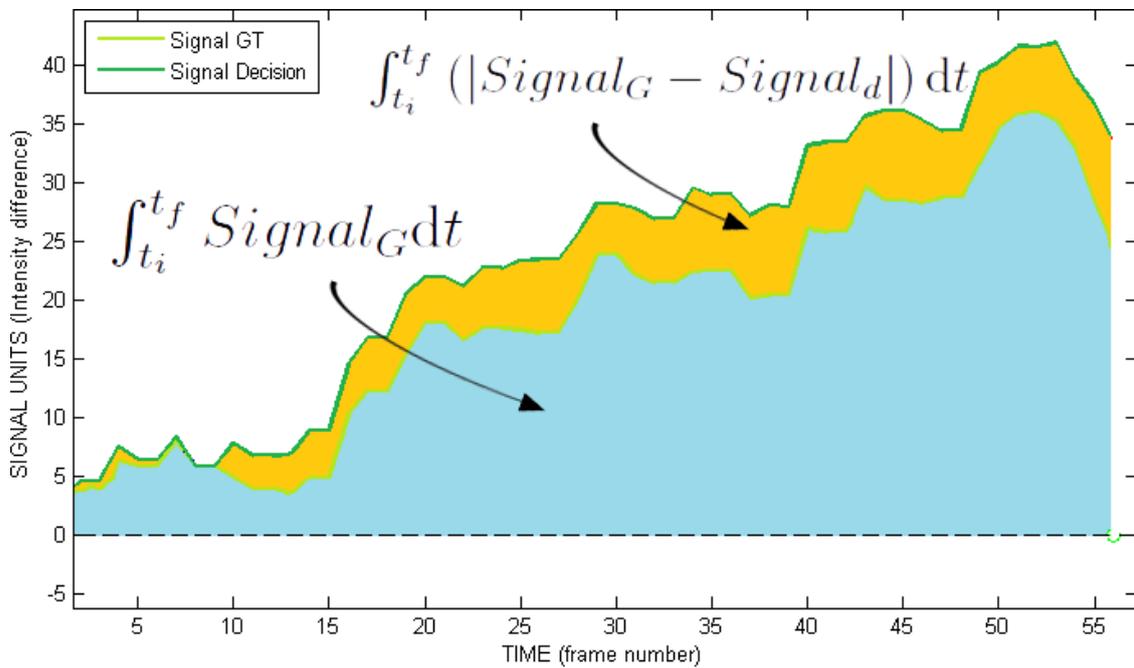


Figure 5. Example of the Normalised Cumulative Signal Difference (NCSD). NCSD is the yellow area divided by the pale blue area.

Finally, a validation at the level of the radiological decision is performed, to assess the final automated decision in comparison with that of a human expert and confirm the usability of the proposed method as a second-opinion tool. This is a binary classification of the FLLs as either benign or malignant, based on their perfusion dynamics and quantified using the Accuracy metric, i.e., the proportion of true results in the whole population.

Evaluation of Degrees of Freedom

To evaluate how many degrees of freedom (DoFs) are needed to sufficiently describe the transformations occurred due to observed motion within the CEUS image plane and therefore obtain sufficiently accurate FLL localizations, the Generalized Procrustes Analysis (GPA) (Gower 1975) has been used in all provided cases.

Firstly, only the cases with the provided non-rigid GT were evaluated. $F_{GT}(t_0)$ has been used as the reference shape, and the $F_{GT}(t)$ of all the frames have been aligned with $F_{GT}(t_0)$ using different transformations, i.e., translation (T), Euclidean (TR) and Similarity (TRS). Specifically, all $F_{GT}(t)$ contours have been sub-sampled to a specified number of points, i.e., half the minimum number of points across all contours. Then, for the transformation of T, the center of gravity (CoG) of each contour $F_{GT}(t)$ was translated to the origin and the overlap of each with $F_{GT}(t_0)$ was assessed. Furthermore, for the TR and the TRS transformations, the minimization of the sum of square distances of the points of each $F_{GT}(t)$ with the points of $F_{GT}(t_0)$ was employed through GPA to find the best fit.

A second experiment is performed on all the provided cases to evaluate the number of DoFs only at the decision level. $F_{GT}(t)$ and $P_{GT}(t)$ are tracked using the same PBRT (i.e., CARD) with different DoFs, producing contours of the FLL and parenchyma for all the frames of the sequence. Then, after using the best mean shape of the FLL, as estimated by the GPA, in the late phase image, the final decision of the FLL's category by the proposed method is compared with the RS.

Finding the Optimal PBRT for the Modality of CEUS

The framework described in the Section "Tracking FLL Contours in CEUS Cine-loops" has been tested with the different PBRTs (Ambai and Yoshida 2011, Bay, et al. 2006, Dalal and Triggs 2005, Evans 2009, Harris and Stephens 1988, Lowe 2004, Rosten and Drummond 2005, Rosten and Drummond 2006, Rublee, et al. 2011, Shi and Tomasi 1994, Vedaldi and Fulkerson 2010), described in Section "Existing Point-Based Registration Techniques (PBRTs)", to find the optimal one for CEUS. For this evaluation, only the pixel level metrics are considered, as we are mainly interested in the correct localization of the FLL at each frame, given its $F_{GT}(t)$ at one of these frames. To assess the saliency and the descriptiveness of the "interest points" detected in the CEUS data, and driven by the results of the section "Evaluation of Degrees

of Freedom”, a translation transformation is considered. Furthermore, the computational efficiency (i.e., the elapsed time) for each PBRT is estimated to assess its usability for real clinical application.

Firstly, the elapsed time required for each PBRT method to run is used to calculate the number of frames processed per second by each of the applied PBRTs as a measure of the efficiency of each one. The numbers shown in Fig.6 are obtained after running the method described in Section “Tracking FLL Contours in CEUS Cine-loops” with each of the different PBRTs (S.Table 1) overall the cases of the GR Dataset. Only the GR Dataset was used due to the expectation that most PBRTs will give significantly low performance and therefore only the ones with an acceptable performance allowing for potential clinical application will be evaluated in all the cases.

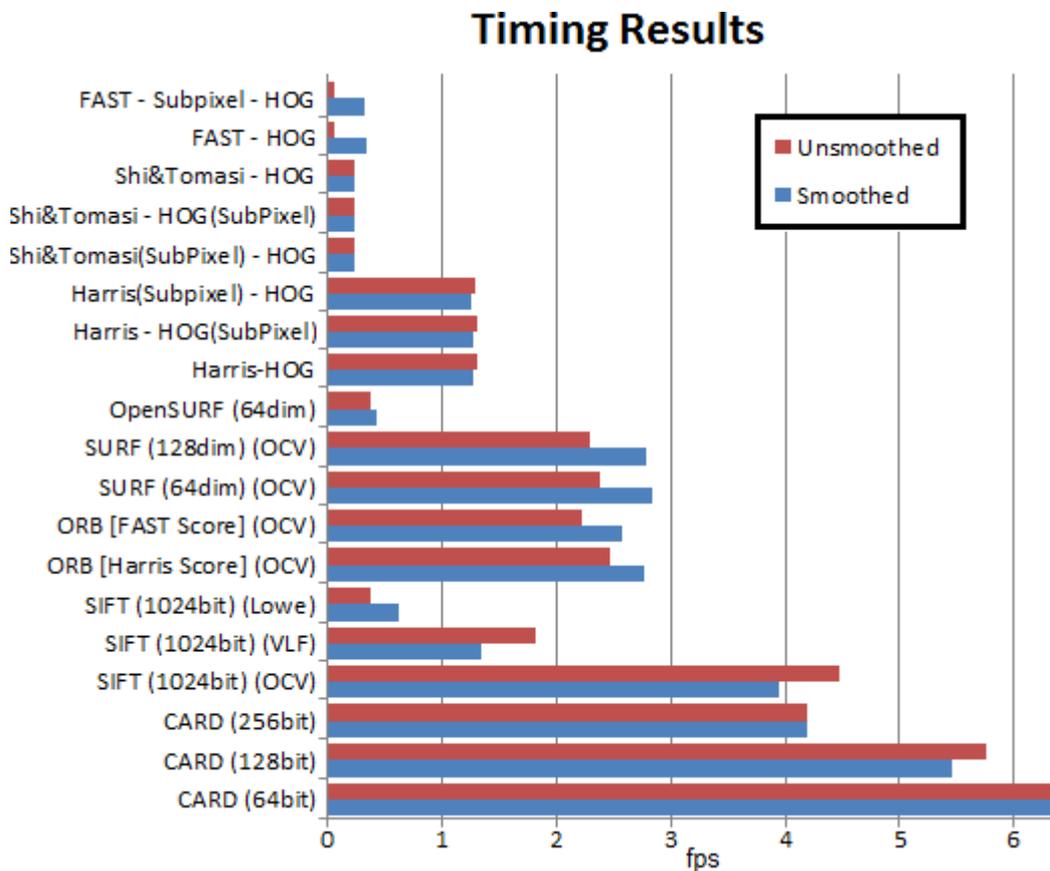


Figure 6. Frames processed per second for different PBRTs, over the total number of frames of the GR cases. OCV and VLF stand for OpenCV and VLFeat libraries, respectively.

Furthermore, the effectiveness of each PBRT is quantitatively evaluated at the pixel level.

Specifically, smoothed and unsmoothed CEUS cine-loops of all the GR cases are evaluated using the Jaccard

and Precision metrics for the FLL region (Fig.7). Note that the Recall results are omitted, as they are identical to those for Precision, since the decision shape is identical with the GT shape.

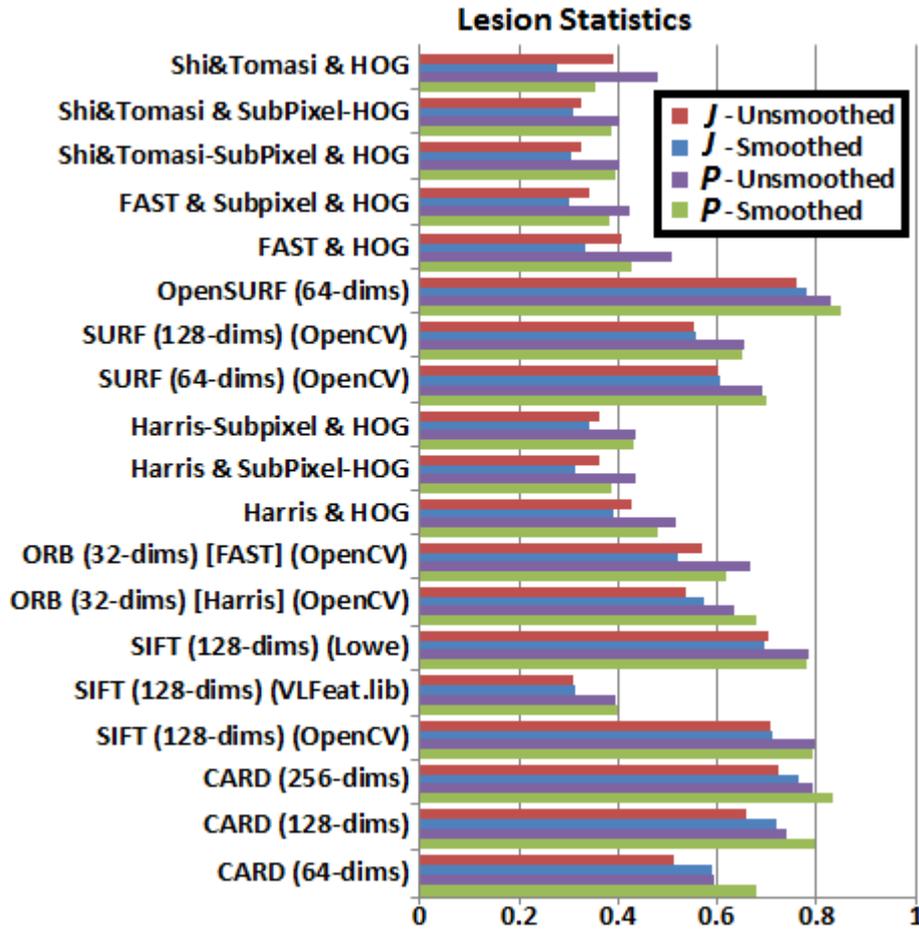


Figure 7. Quantitative evaluation for the FLL region at the pixel level for all the cases of the GR Dataset.

Frame Rate Sampling

The frame rate and the total number of frames in a sequence impacts the time needed for its offline assessment. Considering the potential advantages of real-time processing, here we attempt to identify redundancies in sequentially acquired frames and set the lower bound for the rate of acquisition of “essential” frames.

Noting that the frame rate of the UK Dataset varies, only the GR Dataset was considered that was captured at 25 fps. This frame rate is down-sampled to decide whether 25 fps are essential, or if this number could be reduced, enabling reduction of the time required for processing the sequence. To evaluate this experiment, the NCSD is used as it considers changes throughout the whole duration of a sequence.

$F_{GT}(t)$ for each of the cases is used and a specified number of frames is employed to obtain its Signal' G . This signal slightly differs from the original Signal G due to the different number of frames used to derive it. NCSD is used to compare Signal' G with Signal G and provide an error rate that describes the deviation between these signals.

Results

Evaluation of Degrees of Freedom

Two experiments were performed and results were assessed using evaluation metrics from all levels. The validation results at the pixel, behavioral and decision level of the evaluation of only the cases with the non-rigid GT are summarized in Table 1, using median values, with the best of them being in bold. Decision level results are identical across the different number of DoFs, due to the small number of cases used. Results for the evaluation of the different DoFs on all cases are provided in Table 2.

Table 1. Results on the number of DoFs useful for CEUS data using all cases (n=11) with non-rigid GT. The provided numbers are the median values of results at pixel and behavioral level metrics. Both indicate that use of 2 DoFs produce better results than the use of 3 (or even 4) DoFs.

Validation Level	Metric	Transformations		
		2DoF, T	3DoF, T+R	4 DoF, T+R+S
Pixel	Precision (P)	0.9441	0.9423	0.9178
	Recall (R)	0.9698	0.9508	0.9341
	Jaccard (J)	0.8709	0.8576	0.8702
	F1 score	0.9491	0.9426	0.9301
Behavioral	MSDR	0.1743	0.1593	0.2889
	NCSD	0.0473	0.0634	0.0532
Decision	Decision/GT	1	1	

Table 2. Results on the number of DoFs useful for CEUS data. Evaluation of different DoFs at the decision level considering all available clinical cases and evaluating only with the RS.

Validation Level	Metric	Transformations		
		2DoF, T	3DoF, T+R	4 DoF, T+R+S
Decision	Accuracy	92%	92%	92%

Finding the Optimal PBRT for the Modality of CEUS

The top-ranked PBRTs were CARD and SIFT (OpenCV implementation) that can process at least 4 and up to 6.5 fps, respectively, whilst the next best performing PBRTs (i.e., ORB and SURF OpenCV implementations) are able to process approximately 2.5 fps that is between 37.5% and 61.5% slower than the best performing PBRT (Fig.6).

From the pixel level evaluation (Fig.7), we note a very similar performance between CARD (128- and 256-bit), the OpenCV implementation of SIFT (128-dims) and OpenSURF (64-dims). However, OpenSURF processing speed is just 0.5 fps (Fig.6), which is considered impractical for clinical use and therefore not further evaluated. The other three PBRTs, including CARD (64-bits) (i.e., the last four PBRTs of Fig.6, 7) are included in a further pixel level evaluation in the complete cohort. Therefore, all cases (n=64) with a GT from both datasets were used to evaluate the performance of the methods and the results are shown in Fig.8. After considering the combination of both the elapsed time and the overlap metrics at the pixel level, CARD with a 128-bit descriptor seems to be the most appropriate method for CEUS recordings. Also, it is shown below that analysis of CEUS recordings in real-time requires a frame processing rate of at least 5 fps.

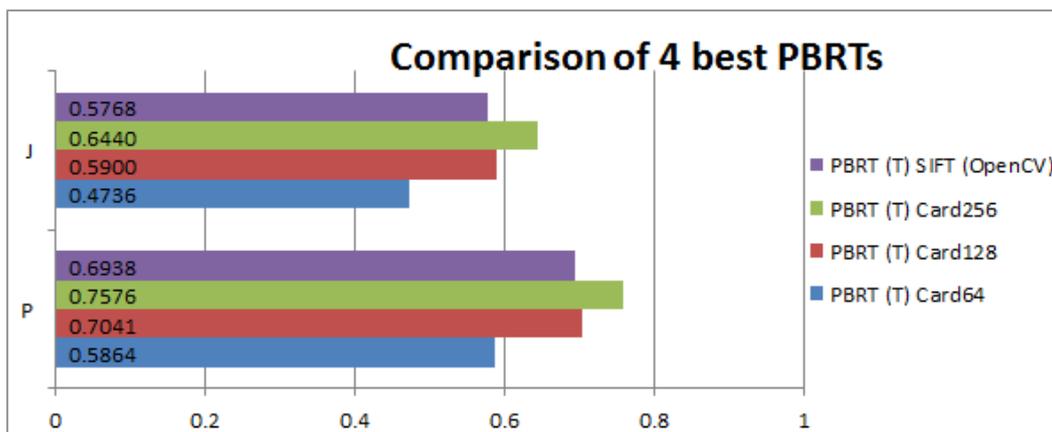


Figure 8. Quantitative evaluation for the FLL region at the pixel level on all available cases, for the four best performing PBRTs, after considering both overlap and run-time performance.

Frame Rate Sampling

As shown in Fig.9, if the sampling rate is less sparse (i.e., greater) than 5 fps, then the median error is below 1%. This, in conjunction with the results of the previous section, supports the applicability of CARD in real-time analysis of CEUS cine-loops at a frame rate of 5 fps.

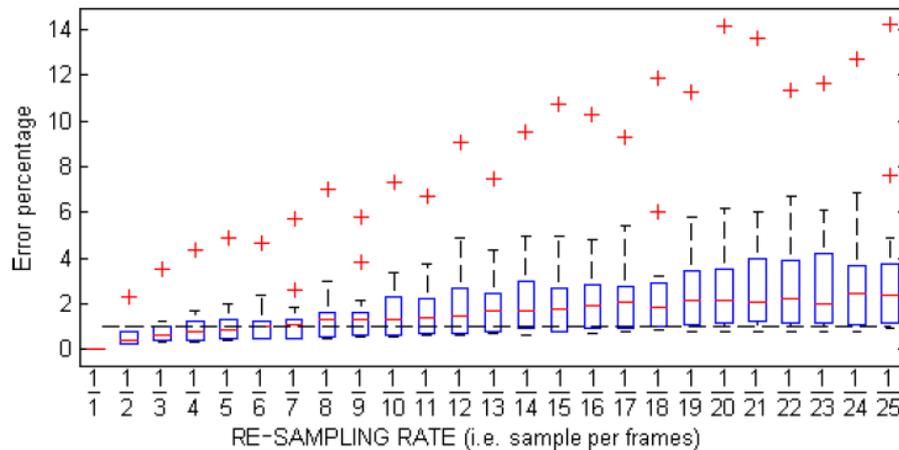


Figure 9. Results on the NCS D for various sampling rates overall cases of the GR Dataset. The values given in the x axis represent the number of samples (numerator) considered per the number of frames specified in the denominator. Note that all these have been acquired at a frame rate of 25 fps.

Discussion

This study focused on the systematic application and evaluation of nineteen configurations and implementations of existing *indirect* methods (i.e., PBRTs) for motion compensation in CEUS cine-loops, and introduced a novel multi-level quantitative evaluation framework, to facilitate the process of quantification of FLLs in CEUS cine-loop obtained using standard clinical practice. To the best of our knowledge, the present study is the first to systematically evaluate such a wide range of PBRTs in the context of CEUS. All previous approaches for motion compensation in CEUS (e.g., (Anaye, et al. 2011, Rognin, et al. 2006, Ta, et al. 2012)) appear to have employed *direct* methods, and these tend to give results which are not precisely repeatable or reproducible since they depend on the manual initialisation used. Similarly, previous CEUS studies of FLL assessment and classification which did not employ any motion compensation have either relied on manual annotation of the FLL within each image frame (Goertz, et al.

2010, Lemke, et al. 2004, Shiraishi, et al. 2008), or relied on the US transducer being held static and the patient remaining very still, and breathing in a way to allow the associated motion to be negligible (Huang-Wei, et al. 2006). Such issues limit the applicability of their approaches and require much input from expert radiologists and/or technician-operators. In contrast, our approach aims to only rely on standard (Claudon, et al. 2013) clinical procedures for CEUS data acquisition and to automate many of the “routine” tasks so that only a modest effort is required on the part of the expert clinician.

In this present paper, a multi-level evaluation framework - at each of the levels of individual pixel changes over time, the overall dynamics of perfusion change, and of final diagnosis - has been proposed to assess the performance of *indirect* methods for motion compensation, in terms of both their computational efficiency and clinical applicability for CEUS scans applied to the detection and quantification of FLLs. As noted above, this is the first such evaluation of indirect motion compensation methods within this domain. The potential for successful quantification of FLLs that the applied methods can provide to radiologists, prompting further investigation when appropriate, was also evaluated, as well as their potential clinical benefits via partially automating the FLL diagnosis process, saving the time and effort of expert radiologists whilst maintaining reliability.

The experiments performed show evidence that only considering displacements of the FLL contours in the x and y directions within the 2D image plane (i.e., changes with 2 DoFs – namely translation transformations), without adjusting the shapes of the ROIs, are sufficient to describe the motion observed in the image plane of a CEUS cine-loop. No consideration of rotations, scalings or distortions were required, at least for the data we had available. However, these results are based on data acquired using standard clinical practice (Claudon, et al. 2013) - no additional specific instructions were imparted to the patient or the operator. For example, possible rotation of the US transducer during a scan might make allowing for rotations an essential element of the compensation method.

Our results also showed that real-time analysis of CEUS recordings requires a frame processing rate of at least 5 fps. Of the 19 different variations of PBRTs considered here, 128-bit CARD (Ambai and Yoshida 2011) appears to be the optimal for motion compensation in CEUS. Based on the given multi-institutional cohort dataset, it provided the best results, in terms of the computational time required for

processing a sequence of frames (Fig.6), along with its evaluation at the pixel, behavioral and decision levels (Fig.7). Specifically, CARD can process and analyze CEUS cine-loops sampled at 5 fps in real-time, whilst use of such a reduced frame rate leads an FLL's apparent dynamic behavior to deviate, on average, less than 1% (according to the NCSD metric) from the corresponding behavior within CEUS cine-loops captured at 25 fps.

Although our study is benefited by the use of real clinical data for its evaluation and this can reflect its potential value in the clinical setting, its design is also limited by the provided data. Specifically, our study did not evaluate any metastatic lesions as they were not part of the provided data. Moreover, the available CEUS cine-loops included only a single FLL within the liver of each patient, and therefore the proposed methodology was not evaluated on scans of multiple lesions. This may be considered a limitation of the current form of "tracking" only two annotated ROIs, e.g., the FLL and the parenchyma. However, this can be easily addressed by defining multiple distinct labels to obtain multiple perfusion curves and it is not an actual limitation of the proposed method, but of its software design that asks only for two ROIs and can be easily adapted, subject to available data for evaluation. Furthermore, the ability of PBRTs to allow for 'global' transformation estimation without requiring the manual annotation of an ROI, and instead of "tracking" an initialized ROI the method will be applied in the principle of "registration". This means that the same methodology will first be applied for estimating and compensating the motion in a CEUS cine-loop by registering the successive frames according to the 'global' estimated transformation, and then the user will annotate multiple ROIs with distinct labels to obtain the multiple perfusion curves. It should also be noted that the provided data available to us from the beginning of this study did not include the B-mode, and hence the focus of this study was to conduct the motion compensation analysis using only the CEUS mode. It is expected that inclusion of the B-mode image plane would offer complementary information to the CEUS plane, particularly about the local structure, and hence analysis of the two planes in tandem should improve the motion compensation even further.

Conclusion

In this paper, we have introduced the first extensive application and evaluation of *indirect* motion compensation methods for facilitating the process of identification, quantification and diagnosis (as benign or malignant) of a focal liver lesion (FLL) from a CEUS cine-loop obtained using standard clinical practice (Claudon, et al. 2013), without requiring any unusually specific data acquisition procedures. Previous approaches applied to CEUS for the diagnosis of FLLs have required substantial manual input from an expert radiologist, and our proposed semi-automatic framework should thus reduce effort and time required per case on the part of the clinician. Amongst the various Point-Based Registration Techniques (PBRTs) considered for motion compensation, the 128-bit CARD (Ambai and Yoshida 2011) approach appears to be the most suitable for this type of CEUS application, offering sufficiently rapid processing of frames for real-time processing whilst maintaining reliability with respect to the perfusion dynamics. Furthermore, based on the clinically-acquired data available to us, it has been demonstrated that considering only translation transformations is normally sufficient to describe the apparent motion of the liver and any FLL observed during a CEUS cine-loop, rather requiring more complex transformations in the motion compensation process.

Future directions for the work presented here include a) further evaluation in prospective FLL CEUS cine-loops following the standard of care protocol and including metastases, b) requesting the B-mode image plane recording with the intention of analyzing it in tandem with the CEUS plane, c) automatic initialisation of ROIs (i.e., parenchyma, FLL). Furthermore, upon packaging of the presented algorithm into a software tool, its design will be adapted to allow analysis of multiple lesions/ROIs. In favor of reproducibility and considering the benefit that readily available tools for computer-aided automation could offer in clinical FLL evaluation, the work described in this paper is part of a larger on-going publicly available open source software toolkit repository (https://github.com/sbakas/FLL_CEUS_analysis). The overarching goal of this repository is to potentially offer a useful advisory approach saving time and effort to expert clinicians, while

allowing them to forge their own evaluation, and also use it as a training/educational tool for the reliable quantification of FLLs in CEUS cine-loops.

Acknowledgements

This work was financially supported by a scholarship provided to Spyridon Bakas, by the Faculty of Science, Engineering and Computing, of Kingston University London for research towards the award of Doctor of Philosophy, for a thesis entitled “Computer-Aided Localisation, Segmentation and Quantification of Focal Liver Lesions in Contrast-Enhanced Ultrasound”.

References

- AIUM. Section 7—Discussion of the Mechanical Index and Other Exposure Parameters. *Journal of ultrasound in medicine : official journal of the American Institute of Ultrasound in Medicine* 2000; 19:143-68.
- Althof RJ, Wind MGJ, Dobbins JT. A rapid and automatic image registration algorithm with subpixel accuracy. *IEEE Transactions on Medical Imaging* 1997; 16:308-16.
- Ambai M, Yoshida Y. 2011 CARD: Compact And Real-time Descriptors. *2011 International Conference on Computer Vision*, 97-104.
- Anaye A, Perrenoud G, Rognin N, Arditi M, Mercier L, Frinking P, Ruffieux C, Peetrons P, Meuli R, Meuwly J-Y. Differentiation of Focal Liver Lesions: Usefulness of Parametric Imaging with Contrast-enhanced US. *Radiology* 2011; 261:300-10.
- 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, Autret A, Hoppe A, Galariotis V, Makris D. Localisation and Characterisation of Focal Liver Lesions using Contrast-Enhanced Ultrasonographic Visual Cues. *Medical Image Understanding and Analysis* 2011.
- Bakas S, Chatzimichail K, Hoppe A, Galariotis V, Hunter G, Makris D. Histogram-based Motion Segmentation and Characterisation of Focal Liver Lesions in CEUS. *Annals of the British Machine Vision Association (BMVA)* 2012; 2012a:1-14.
- 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* 2017a; 5:329-38.

- 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) 2012b; 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, Makris D, Hunter GJA, Fang C, Sidhu PS, Chatzimichail K. Automatic Identification of the Optimal Reference Frame for Segmentation and Quantification of Focal Liver Lesions in Contrast-Enhanced Ultrasound. *Ultrasound in Medicine & Biology* 2017b; 43:2438-51.
- Bay H, Tuytelaars T, Van Gool L. SURF: Speeded Up Robust Features, In: Leonardis A, Bischof H, and Pinz A, eds. *Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. 404-17.
- Bradski G. The OpenCV Library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer* 2000; 25:120-23.
- Burrowes DP, Medellin A, Harris AC, Milot L, Wilson SR. Contrast-enhanced US Approach to the Diagnosis of Focal Liver Masses. *RadioGraphics* 2017; 37:1388-400.
- Calonder M, Lepetit V, Strecha C, Fua P. BRIEF: Binary Robust Independent Elementary Features, In: Daniilidis K, Maragos P, and Paragios N, eds. *Computer Vision – ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part IV*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. 778-92.
- Canny J. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1986; PAMI-8:679-98.
- Castro ED, Morandi C. Registration of Translated and Rotated Images Using Finite Fourier Transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1987; PAMI-9:700-03.
- 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, Correias 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.
- Correas JM. 2014 CEUS and New Tools for Percutaneous Interventional Procedures. *European Congress in Radiology (ECR)*. Austria, Vienna.
- Dalal N, Triggs B. 2005 Histograms of oriented gradients for human detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 886-93 vol. 1.
- Derek LGH, Philipp GB, Mark H, David JH. Medical image registration. *Physics in Medicine & Biology* 2001; 46:R1.
- Dietrich CF, Ignee A, Hocke M, Schreiber-Dietrich D, Greis C. Pitfalls And Artefacts Using Contrast Enhanced Ultrasound. *Z Gastroenterol* 2011; 49:350-56.
- Dill-Macky MJ, Burns PN, Khalili K, Wilson SR. Focal Hepatic Masses: Enhancement Patterns with SH U 508A and Pulse-Inversion US. *Radiology* 2002; 222:95-102.
- 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.
- Ding L, Goshtasby A, Satter M. Volume image registration by template matching. *Image and Vision Computing* 2001; 19:821-32.
- Evans C. Notes on the opensurf library. University of Bristol, Tech. Rep. CSTR-09-001, January 2009.
- Fitzpatrick JM, West JB. The distribution of target registration error in rigid-body point-based registration. *IEEE Transactions on Medical Imaging* 2001; 20:917-27.

- Goertz RS, Bernatik T, Strobel D, Hahn EG, Haendl T. Software-based quantification of contrast-enhanced ultrasound in focal liver lesions—A feasibility study. *European Journal of Radiology* 2010; 75:e22-e26.
- Gower JC. Generalized procrustes analysis. *Psychometrika* 1975; 40:33-51.
- Harris C, Stephens M. 1988 A combined corner and edge detector. *Alvey vision conference: Manchester, UK*, 10.5244.
- Harvey CJ, Blomley MJK, Eckersley RJ, Cosgrove DO. Developments in Ultrasound Contrast Media. *Eur Radiol* 2001; 11:675–89.
- Huang-Wei C, Bleuzen A, Bourlier P, Roumy J, Bouakaz A, Pourcelot L, Tranquart F. Differential Diagnosis of Focal Nodular Hyperplasia With Quantitative Parametric Analysis in Contrast-Enhanced Sonography. *Investigative Radiology* 2006; 41:363-68.
- Irani M, Anandan P. About Direct Methods, In: Triggs B, Zisserman A, and Szeliski R, eds. *Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms Corfu, Greece, September 21–22, 1999 Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000. 267-77.
- Khajone BB, Kokate RB, Shandilya V. A Survey of Image Registration Techniques. *IJRIT International Journal of Research in Information Technology* 2014; 2:554-60.
- Lemke AJ, Chopra SS, Hengst SA, Brinkmann MJ, Steinmüller T, Felix R. [Characterization of hepatic tumors with contrast-enhanced ultrasound and digital grey-scale analysis]. *Rofo* 2004; 176:1607-16.
- Leutenegger S, Chli M, Siegwart RY. 2011 BRISK: Binary Robust invariant scalable keypoints. *2011 International Conference on Computer Vision*, 2548-55.
- Lowe DG. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 2004; 60:91-110.
- Lu X, Zhang S, Su H, Chen Y. Mutual information-based multimodal image registration using a novel joint histogram estimation. *Computerized Medical Imaging and Graphics* 2008; 32:202-09.
- Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging* 1997; 16:187-98.

- Makela T, Clarysse P, Sipila O, Pauna N, Quoc Cuong P, Katila T, Magnin IE. A review of cardiac image registration methods. *IEEE Transactions on Medical Imaging* 2002; 21:1011-21.
- Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2005; 27:1615-30.
- Norouzi M, Fleet DJ, Salakhutdinov RR. 2012 Hamming distance metric learning. *Advances in neural information processing systems*, 1061-69.
- Penrose R. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society* 2008; 51:406-13.
- Piccardi M. 2004 Background subtraction techniques: a review. *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, 3099-104 vol.4.
- Pluim JPW, Maintz JBA, Viergever MA. Mutual information matching in multiresolution contexts. *Image and Vision Computing* 2001; 19:45-52.
- Rognin N, Campos R, Thiran J, Messenger 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.
- Rosten E, Drummond T. 2005 Fusing points and lines for high performance tracking. *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, 1508-15 Vol. 2.
- Rosten E, Drummond T. Machine Learning for High-Speed Corner Detection, In: Leonardis A, Bischof H, and Pinz A, eds. *Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. 430-43.
- Rothwell CA, Zisserman A, Forsyth DA, Mundy JL. Planar object recognition using projective shape representation. *International Journal of Computer Vision* 1995; 16:57-99.
- Rublee E, Rabaud V, Konolige K, Bradski G. 2011 ORB: An efficient alternative to SIFT or SURF. *2011 International Conference on Computer Vision*, 2564-71.
- Schmid C, Mohr R, Bauckhage C. Evaluation of Interest Point Detectors. *International Journal of Computer Vision* 2000; 37:151-72.

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

Schneider M, Arditi M, Barrau M, Brochot J, Broillet A, Ventrone R, Yan F. BR1: a new ultrasonographic contrast agent based on sulfur hexafluoride-filled microbubbles. *Invest Radiol.* 1995; 30:451-7.

Shekhar R, Zagrodsky V. Mutual information-based rigid and nonrigid registration of ultrasound volumes. *IEEE Transactions on Medical Imaging* 2002; 21:9-22.

Shi J, Tomasi C. 1994 Good features to track. *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 593-600.

Shiraishi J, Sugimoto K, Moriyasu F, Kamiyama N, Doi K. Computer-Aided Diagnosis for the classification of Focal Liver Lesions by Use of Contrast-Enhanced Ultrasonography. *Medical Physics* 2008; 35:1734-46.

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.

Sotiras A, Davatzikos C, Paragios N. Deformable medical image registration: a survey. *IEEE Trans Med Imaging* 2013; 32:1153-90.

Ta CN, Kono Y, Barback CV, Mattrey RF, Kummel AC. Automating tumor classification with pixel-by-pixel contrast-enhanced ultrasound perfusion kinetics. *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena* 2012; 30:02C103.

Tang M-X, Mulvana H, Gauthier T, Lim AKP, Cosgrove DO, Eckersley RJ, Stride E. Quantitative contrast-enhanced ultrasound imaging: a review of sources of variability. *Interface Focus* 2011; 1:520-39.

Tiemann K, Lohmeier S, Kuntz S, Koster J, Pohl C, Burns P, Porter TR, Nanda NC, Luderitz B, Becher H. Real-Time Contrast Echo Assessment of Myocardial Perfusion at Low Emission Power: First Experimental and Clinical Results Using Power Pulse Inversion Imaging. *Echocardiography* 1999; 16:799-809.

Torr PH, Zisserman A. 1999 Feature based methods for structure and motion estimation. *Workshop on vision algorithms*: Springer, 278-94.

- Vedaldi A, Fulkerson B. 2010 Vlfeat: an open and portable library of computer vision algorithms. *Proceedings of the 18th ACM international conference on Multimedia*. Firenze, Italy: ACM, 1469-72.
- Viola P, Jones MJ. Robust Real-Time Face Detection. *International Journal of Computer Vision* 2004; 57:137-54.
- Wilson SR, Burns PN. Microbubble-Enhanced US in Body Imaging: What Role? *Radiology* 2010; 257:24–39.
- Wilson SR, Lyshchik A, Piscaglia F, Cosgrove D, Jang H-J, Sirlin C, Dietrich CF, Kim TK, Willmann JK, Kono Y. CEUS LI-RADS: algorithm, implementation, and key differences from CT/MRI. *Abdominal Radiology* 2018; 43:127-42.
- Wollny G, Kellman P, Santos A, Ledesma-Carbayo MJ. Automatic motion compensation of free breathing acquired myocardial perfusion data by using independent component analysis. *Medical Image Analysis* 2012; 16:1015-28.

Supplementary Material

Indirect Method (PBRT)			Statistical Descriptor	Implementation	Distance Metric
Geometric Feature Detector		Name			
Name	Implementation	Name	Implementation		
CARD (Ambai and Yoshida 2011)	Authors	CARD 64-bits (Ambai and Yoshida 2011)	Authors	Hamming Distance	
CARD (Ambai and Yoshida 2011)	Authors	CARD 128-bits (Ambai and Yoshida 2011)	Authors	Hamming Distance	
CARD (Ambai and Yoshida 2011)	Authors	CARD 256-bits (Ambai and Yoshida 2011)	Authors	Hamming Distance	
FAST (Rosten and Drummond 2005)	OpenCV	HOG (Dalal and Triggs 2005)	OpenCV	Euclidean Distance	
FAST (Rosten and Drummond 2005) + SubPixel	OpenCV	HOG (Dalal and Triggs 2005)	OpenCV	Euclidean Distance	
Harris (Harris and Stephens 1988)	MATLAB	HOG (Dalal and Triggs 2005)	OpenCV	Euclidean Distance	
Harris (Harris and Stephens 1988) + SubPixel	MATLAB	HOG (Dalal and Triggs 2005)	OpenCV	Euclidean Distance	
Harris (Harris and Stephens 1988)	MATLAB	HOG (Dalal and Triggs 2005) + SubPixel	OpenCV	Euclidean Distance	
ORB [FAST Score] (Rublee, et al. 2011)	OpenCV	ORB 32-dims	OpenCV	Euclidean Distance	
ORB [Harris Score] (Rublee, et al. 2011)	OpenCV	ORB 32-dims	OpenCV	Euclidean Distance	
Shi & Tomasi (Shi and Tomasi 1994)	MATLAB	HOG (Dalal and Triggs 2005)	OpenCV	Euclidean Distance	
Shi & Tomasi (Shi and Tomasi 1994) + SubPixel	MATLAB	HOG (Dalal and Triggs 2005)	OpenCV	Euclidean Distance	
Shi & Tomasi (Shi and Tomasi 1994)	MATLAB	HOG (Dalal and Triggs 2005) + SubPixel	OpenCV	Euclidean Distance	
SIFT (Lowe 2004)	Authors	SIFT 128-dims (1024-bits)	Authors	Ratio of Angles	
SIFT (Lowe 2004)	OpenCV	SIFT 128-dims (1024-bits)	OpenCV	Euclidean Distance	
SIFT (Lowe 2004)	VLFeat	SIFT 128-dims (1024-bits)	VLFeat	L2 norm (VLFeat Implem.)	
SURF (Bay, et al. 2006)	OpenCV	SURF 64-dims (Bay, et al. 2006)	OpenCV	Ratio of Angles	
SURF (Bay, et al. 2006)	OpenCV	SURF 128-dims (Bay, et al. 2006)	OpenCV	Ratio of Angles	
SURF (Bay, et al. 2006)	OpenSURF	SURF 64-dims (Bay, et al. 2006)	OpenSURF	Ratio of Angles	

Supplementary Table 1. Configurations of PBRTs used in this study (in alphabetical order), denoting the descriptors, the distance metric, and the specific implementation used. Note that the matching strategy used for each PBRT was the Nearest Neighbor Distance Ratio (NNDR).