Automatic Identification and Localisation of Potential Malignancies in Contrast-Enhanced Ultrasound Liver Scans Using Spatio-Temporal Features

Spyridon Bakas[1], Dimitrios Makris[1], Paul S.Sidhu[2], Katerina Chatzimichail[3]

[1] Digital Imaging Research Centre, Faculty of Science, Engineering and Computing, Kingston University, London, UK.
`{<S.Bakas, D.Makris> @kingston.ac.uk}`
[2] King's College Hospital, Department of Radiology, Denmark Hill, London, UK.
`{PaulSidhu@nhs.net}`
[3] Evgenidion Hospital, National & Kapodistrian University, Athens, Greece.
`{Katerina@hcsl.com}`

Abstract. The identification and localisation of a focal liver lesion (FLL) in Contrast-Enhanced Ultrasound (CEUS) video sequences is crucial for liver cancer diagnosis, treatment planning and follow-up management. Currently, localisation and classification of FLLs between benign and malignant cases in CEUS are routinely performed manually by radiologists, in order to proceed with making a diagnosis, leading to subjective results, prone to misinterpretation and human error. This paper describes a methodology to assist clinicians who regularly perform these tasks, by discharging benign FLL cases and localise potential malignancies in a fully automatic manner by exploiting the perfusion dynamics of a CEUS video. The proposed framework uses local variations of intensity to distinguish between hyper- and hypo-enhancing regions and then analyse their spatial configuration to identify potentially malignant cases. Automatic localisation of the potential malignancy on the image plane is then addressed by clustering, using Expectation-Maximisation for Gaussian Mixture Models. A novel feature that combines description of local dynamic behaviour with spatial proximity is used in this process. Quantitative evaluation, on real clinical data from a retrospective multi-centre study, demonstrates the value of the proposed method.

Keywords: Localisation, Malignancy Identification, Contrast-Enhanced Ultrasound, Focal Liver Lesion, Liver, Perfusion, Clustering.

1 Introduction

Contrast-Enhanced Ultrasound (CEUS) is a modality widely accepted for the detection and characterisation of focal liver lesions (FLLs)[6]. In comparison with conventional B-mode ultrasound (US), CEUS enhances the contrast between the liver and the FLL through the use of intravenously injected con-

trast agents [9]. CEUS has diagnostic accuracy higher than 95% for the evaluation of malignant FLLs [16] and its sensitivity and specificity exceeds that of other modalities, such as contrast-enhanced computed tomography (CE-CT) and contrast-enhanced magnetic resonance imaging (CE-MRI) [18]. It is recognised as the most cost-efficient imaging solution for distinguishing between benign and malignant FLLs [6], after an inconclusive US scan. Also it is in the forefront of CEUS scope to help in reducing the radiation burden to population.

A CEUS liver scan is divided into three different phases over time, namely arterial, portal venous and late phase. These allow the observation of the flow of the intravenously injected contrast agent, by intensity changes in the captured plane. Recording these intensity changes for different tissues during a CEUS sequence leads to time-intensity curves (TICs). These curves describe the perfusion dynamics for different regions and lead to parameters, that allow for the differentiation of the nature of tissues [6]. Specifically, a tissue that shows increased perfusion (hyper-enhancing) in comparison with the rest healthy tissue (parenchyma), during the arterial phase, reveals the typical behaviour of a potentially malignant FLL. This FLL can only be confirmed as a malignancy if during the late phase its region is darker than the parenchyma [19].

Currently, radiologists routinely detect, localise and classify FLLs in CEUS data manually, through a time-consuming series of tasks, namely **(i)** identification of a reference frame for **(ii)** localising a region of interest (ROI), e.g. FLL, **(iii)** monitoring of the dynamic behaviour (i.e. brightness intensity changes) of different ROIs over time, and eventually classifying an FLL as benign or malignant. Each of these tasks requires a high-level of expertise, provides subjective results and is prone to misinterpretation and human error [6]. Processing of CEUS data poses a very challenging task due to intensity changes, acoustic shadows, low signal-to-noise ratio, the transducer movement, as well as the patient's irregular breathing patterns and the motion of the inner human organs that affect the ROIs' apparent 2D size and shapes.

While computer-aided solutions have been suggested for the aforementioned tasks of **(i)** [4] and **(iii)** [2] [3] [5] [12] [17] in CEUS data, no solution has ever been proposed for the automated localisation of the position and shape of an FLL in a CEUS recording (task **(ii)**). Considering the continuously increasing number of CEUS scans, automation of the localisation procedure is of much interest as it will assist the diagnostic procedure and the assessment of the repeatability and reproducibility of the examination on different patients or between different clinical centres, as it is expected to provide objective results.

Methods suggested thus far for the task of monitoring FLLs in CEUS (task **iii**) require the existence of prior manually initialised ROIs (e.g. FLL). A method is described in [12], which first compensates for rigid motion in CEUS video recordings and then displays TICs to assist radiologists to classify an FLL. Motion compensation is based on iterative maximisation of mutual information [14], which is dependent on an intensity constancy constraint. Another method described in [2], creates a non-rigid motion model by combining histogram information from a ROI with the detection of SIFT keypoints [10]. It then suggests

an FLL classification decision based on obtained TICs. Both methods require the manual delineation of the ROI and the part of the image that needs re-alignment (e.g. the conical area of the ultrasonographic image, also named US mask) on a manually identified frame. Then another method, suggested in [3], proposes an affine transformation motion model, based entirely on SIFT key-points [10], to monitor and classify an FLL based on its TIC. In addition, [3] introduces an automatic delimitation of the US mask based on an intensity change detection. A pixelwise analysis within a manually initialised FLL is performed in [17]. Specifically, after compensating for rigid motion similar to [12], micro-vasculature differences between benign and malignant FLLs are compared and the FLLs' spatial heterogeneity is quantified.

Furthermore, segmentation methods for the localisation of lesions have been applied in CT [20] and MRI [7] [15] data. Specifically, these methods require some manually annotated areas in one slice by an operator, and then they either encode the pairwise similarity of intensity information between different pixels, or model the intensity relationship of each pixel to a set of global intensity clusters, across the 3rd dimension of the data (i.e. different depth slice). Even though the 3rd dimension in volumetric CT and MRI data can be considered relative to the temporal dimension of CEUS data, in CT and MRI the same tissues are represented with consistent brightness intensity across the 3rd dimension, making the localisation task much easier, especially if prior knowledge is included.

The aim of the present study is firstly to identify and categorise hyper- and hypo-enhancing FLLs in the first phase of a CEUS examination, and then to localise regions of potential malignancy within the liver region in CEUS video sequences, in a fully automatic manner and without considering any prior knowledge. This can be ultimately used as an objective initialisation of segmentation methods, as well as to attract the attention of radiologists in the suggested ROIs within the US mask that they might have missed.

2   Method

Specifically, this paper considers the problem of identifying and localising a potential malignancy by clustering together regions with the same dynamic behaviour (TIC) within the US mask, in a similar manner to the process performed conceptually by radiologists. After compensating for the motion included within the sequence of a patient's scan, the US mask is divided into smaller regions and their local TICs are obtained. Subsequently, dimensionality reduction is applied on the TICs for each individual patient, and the principal axes describing more than 90 % of the total variance are used in combination with location information to form the feature vector for each region. A framework based on the optimisation technique of Expectation-Maximisation (EM) with a set of Gaussian Mixture Models (GMM) is then used to segment the US mask into meaningful ROIs based on these feature vectors, and therefore localise the hyper-enhancing regions, i.e. potential malignancies. Fig. 1 summarises the whole pipeline of the proposed method.
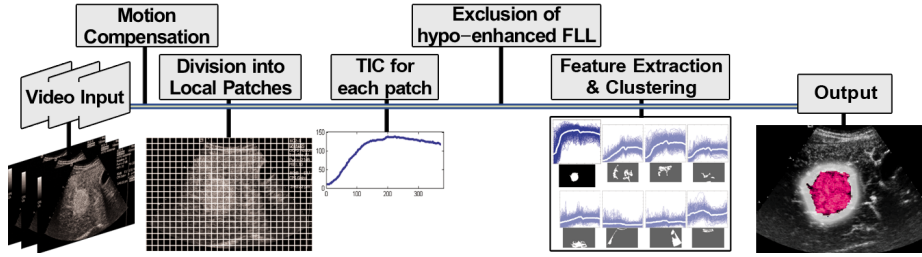
Fig. 1. Visual summary of the proposed method.

Initially, the US mask is automatically segmented on each patient's case by considering an intensity change detection, as described in [3], in order to remove all the irrelevant information - notably textual data provided for the CEUS operator. This US mask is then applied to every frame of the sequence, selecting only the relevant information, whilst removing the aforementioned artefacts. Then the optimal reference frame at time t_0 is automatically identified, as proposed in [4]. According to radiologists, this frame is expected to be the one with the maximum contrast between FLL and parenchyma.

Any present motion between the liver and the US transducer affects negatively the processing outcome and therefore the frames of each sequence need to be spatially realigned to compensate for any in-plane movements. Assuming there is only motion within –instead of across– the plane, and a simple translation is sufficient to describe the level of relative motion between the liver and the US transducer, the point-based registration technique of Compact And Real-time Descriptors (CARD) keypoints [1] is employed to automatically estimate the motion between all the frames and realign them according to the reference frame. This is done by first registering keypoints every two successive frames of the acquired sequence and then matching correspondent keypoints by using the Nearest Neighbour Distance Ratio in the descriptor space. The translation between every pair of successive frames is then estimated as the average displacement of the matched correspondences between them. The motion compensated video data can then be processed as a 3D volume, where the 3rd dimension depicts information from the temporal domain.

2.1 Local Time-Intensity Curves

Obtaining information from the brightness intensity of a single pixel (fine-grained resolution) is considered susceptible to speckle noise (i.e. very sensitive to "outliers" – excessively bright and dark pixels). Therefore, the spatial averaging through more coarse-grained resolutions was considered essential, by subdividing the US mask into B non-overlapping local neighbourhoods ("patches") of $n \times n$ pixels and avoid such an effect, where n is small compared with the over-

all size of the image. The pixel brightness intensity values are then averaged over each of these patches.

$$\bar{p}_b(t) = \frac{1}{n^2} \sum_{i_b=1}^{n^2} p_{i_b}(t). \tag{1}$$

where $p_{i_b}(t)$ is the intensity of the ith pixel within the bth block at time t.

The changes of the average brightness intensities of each patch ($\bar{p}_b$) over time (i.e its dynamic behaviour) can lead to information similar to that used by radiologists. Such information are encompassed in the TICs of the local patches, by estimating the $\bar{p}_b$ for each frame in the sequence.

$$\mathbf{TIC_b} = [\bar{p}_b(1),\, \bar{p}_b(2),\, \ldots,\, \bar{p}_b(T)]. \tag{2}$$

where T refers to the number of frames of the CEUS video sequence.

2.2   Early Exclusion of Hypo-enhancement

A foreground mask is produced including only patches where potential malignancies may exist. Specifically, a patch is included in the mask if it is enhancing more than a certain proportion (ψ) of the maximum intensity value of the whole 3D volume. In addition, a morphological opening is used as noise removal, to exclude any patches that are not connected with any other neighbouring patches, or even if connected, the area that they cover is smaller than the size (ϕ) of the structuring element used for the morphological opening.

The resulting foreground mask for a hyper-enhancing case is expected to be solid, as the FLL would have enriched as much as the parenchyma until the end of the arterial phase. On the other hand, for a hypo-enhancing case, this mask is expected to have a 'hole' either in the middle or at the side of the mask, depending on where the FLL is located in relation to the parenchyma. This property is exploited to automatically discharge hypo-enhancing cases by separating them from hyper-enhancing cases after applying a specified threshold on the overlap metric of the Jaccard index between the foreground mask and its convex hull.

2.3   Feature Extraction and Clustering

Information of the dynamic behaviour of the foreground patches is essential to be considered in the clustering of the data. However, because T is large (see Section 3.1), the data included in each TIC is represented as a high-dimensional vector. Thus, principal component analysis (PCA) by eigenvalue decomposition is employed to reduce the dimensionality of the data and the computational complexity, as well as to create more meaningful feature vectors. The number of dimensions (d) is chosen such that the variance in the data accounted for by these d dimensions (v_d), exceeds a specified proportion, typically 90%, of the total variance v_T ($v_d > 0.9 \cdot v_T$, for $1 \leq d \leq T$). The coordinates of the centre
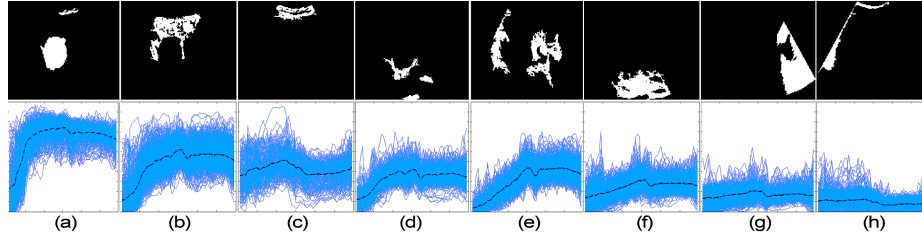
Fig. 2. Clusters obtained from a clinical case, during arterial phase, sorted in descending order of their regional blood flow. *The first row depicts the clusters visualised in the image space and the second row depicts the corresponding TICs, where the vertical and horizontal axes denote brightness intensity and time, respectively. Each of the curves depicts a different patch and the dark dotted curve in the middle of each of the 8 graphs is the average/centroid for each cluster.*

of each patch (x_{c_b}, y_{c_b}) are also included in each feature vector along with the PCA dimensions of the local TIC, such that each feature vector may represent both the spatial proximity and the dynamic behaviour of the local patch.

The optimisation technique of Expectation-Maximisation (EM) is used with a set of Gaussian Mixture Models (GMM) to perform the clustering of the feature vectors. This is done in an attempt to cluster together patches with similar dynamic behaviour. To find the optimal number of clusters for each case, an overestimated initial number of clusters is used to initialise a Bayesian Classifier with the functionality of a GMM Probability Distribution Function, based on the Figueiredo-Jain algorithm [8].

2.4   Region Selection

Radiologists extract various parameters from a TIC to describe different aspects of a ROI's perfusion and determine its functional features [6]. However, most of these parameters require a continuous sequence throughout the three phases of the exam. Due to the provided data comprising only a short sequence of the exam, including at most a small part of the portal venous phase, the only parameters possible to extract are: the peak intensity (PI), the time to reach the PI (TPI) and the regional blood flow (RBF), which is the area under the TIC from time zero until TPI. RBF is considered as the most useful parameter for the scope of this work as it implicitly includes information about the PI, the TPI and the slope of the TIC.

Sorting the identified clusters in descending order of their RBF is expected to reveal first the cluster with the most hyper-enhancing behaviour. In case of an actual hyper-enhancing FLL this cluster can provide sufficient information for the location, size, shape and nature of the FLL, whilst if it is an actual hypo-enhancing case, then the proposed method is expected to identify regions of the parenchyma that have a hyper-enhancing behaviour in relation to the rest of

the image, if not a second FLL. For example, in Fig. 2, cluster depicted by (a) shows the largest RBF, which is typical of a malignant FLL. Furthermore, as the chosen cluster might include some irrelevant pixels (i.e. artefacts) on the image plane (Fig. 2), a proportion of the cluster's pixel population is used (e.g. the 1st quartile) to automatically provide the location of the potential malignancy on the image plane. This location can ultimately be used as an initialisation seed-point to an existing segmentation method.

3 Experiments and Results

3.1 Materials

The proposed method is evaluated through being applied to real clinical data of 63 case studies, of patients in different physical condition. These cases comprise 34 hyper- and 29 hypo-enhancing FLLs during the arterial phase of a CEUS scan and in most cases information from the portal venous phase was also included. Each case includes one sequence of duration between 3 and 82 seconds. The imaging plane was chosen such that most of the motion is within –instead of across– the plane, allowing for its compensation.

All data was acquired using Siemens ACUSON Ultrasound (US) systems (Mountainview CA). Specifically, 49 cases were captured at King's College Hospital in the UK, using an S2000 US system equipped with 4 (or) 6 MHz curvilinear transducer at spatial resolution 1024×768 pixels, and the remaining 14 cases were captured at Evgenidion Hospital in Greece, using a Sequoia C512 US system equipped with 6-2 MHz curvilinear transducer, at spatial resolution 768×576 pixels. In all examinations the second generation contrast medium SonoVue [13] (Bracco S.p.A., Italy) was used in a 2.4ml bolus intravenous injection (into an arm vein). Specific acquisition parameters of the equipment, such as the transducer's field of view and gain, for each patient are unknown, as they were set by the radiologist individually at the start of each examination. The acquisition method of this data reflects true clinical practice and leads to increased variability. Examinations were performed by radiologists with 13-16 years of experience using CEUS. All data were obtained without prior knowledge of subsequent processing by a software tool and without any specific instructions being given to the radiologist beforehand, hence reflecting true clinical practice. Appropriate ethics and confidentiality procedures have been followed at all times.

3.2 Results

For evaluating the automatic identification of hyper-enhancing (i.e. potentially malignant) cases, the measures of true positive rate (sensitivity) and false positive rate (1-specificity) are used, in the range [0,1]. The positive and the negative samples refer to the hyper- and the hypo-enhancing cases, respectively. From the clinical point of view, the true positive (TP) rate should ideally be kept equal to 1, so no single potentially malignant case is missed. At the same time, the
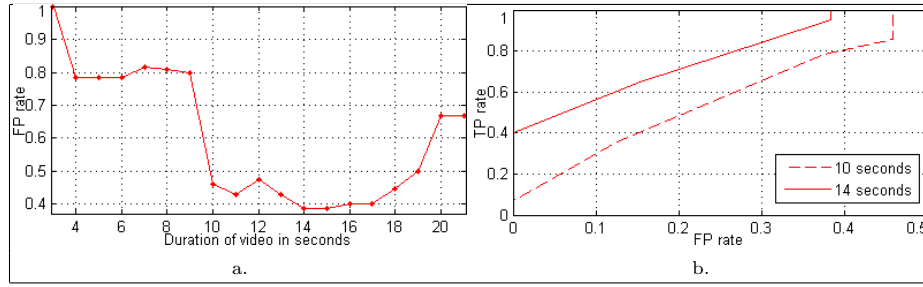
Fig. 3. Graph (a) denotes values of FP rate *(y axis)* for the acquisition duration of sequences being more than a certain number of seconds *(x axis)*, whilst TP rate is equal to 1. Graph (b) shows the ROC curves for cases with duration $\geq$ 10 & 14 seconds.

proposed method should discharge as many hypo-enhancing cases as possible, keeping the false positive (FP) rate low.

According to [6] the duration of the arterial phase is at least 10 seconds. This justifies our results, shown in Fig. 3.a, where the FP rate is as high as 0.8 (specificity=0.2) when sequences with acquisition duration (AD) less than 10 seconds are considered. Then for sequences with AD more than 10 seconds, the FP rate drops to 0.46 (specificity=0.54). The best FP rate is equal to 0.38 (specificity=0.62) and obtained when the AD is at least 14 seconds. Furthermore, two ROC curves are shown in Fig. 3.b. to assess the possible change of the FP rate in relation to the TP rate, for both AD above 10 and 14 seconds.

For evaluating the automatic localisation of the potential malignancies on the image plane, first the boundaries of the FLL have been manually annotated by a radiologist in the reference frame providing its ground truth (GT). The performance of the method is then assessed based on whether the location point provided by the proposed method (described in Section 2.4) is within the GT, or not. To measure this performance a correct localisation rate is used, over the number of hyper-enhancing cases with AD above 10 and 14 seconds, separately.

To provide some comparative results, a baseline approach is considered based on automatically thresholding the reference frame, using the Otsu's method [11]. This is expected to provide a foreground population of pixels that describe mostly the hyper-enhancing FLL. Then, similarly to the proposed method, a proportion of this population is used to obtain the location of the potential malignancy.

The best correct localisation rate achieved is 80% and obtained for the 25% of the chosen cluster's population (Table 1). On the other hand, the best result for the baseline method is 62.07% after using 40% of its mask's population.

4   Conclusions and Future Work

This paper demonstrates that the task of identifying and localising a potentially malignant FLL can be performed in a fully automatic manner. Specifically, the

Table 1. Correct localisation rate for a seed-point within the actual FLL.

Method	Population's proportion	Duration thr≥10"	Duration thr≥14"
Baseline	(25%)	31.03%	25%
Baseline	(40%)	62.07%	55%
Proposed method	(25%)	51.72%	**80%**
Proposed method	(40%)	**65.52%**	75%

proposed framework firstly distinguishes between hyper- and hypo-enhancing cases by using only a video sequence of the arterial phase as input, and then automatically localises potential malignancies on the image plane. The first step is addressed by assessing local intensity variations and analysing their spatial configuration. Then, for the localisation step, a novel feature vector that encompasses the local dynamic behaviour and combines it with the spatial proximity is used in a clustering approach, using EM-GMM.

Experimental results show that the proposed method appears to perform adequately on identifying and localising FLLs with hyper-enhancing behaviour during the arterial phase. Such lesions are of significant importance to radiologists, as they may account for malignancies. FLLs of typical hypo-enhancing behaviour during the arterial phase are benign and therefore of less importance to radiologists.

Further improvements of the proposed framework should include the reduction of the FP rate, making the identification of potential malignancies more specific. Also coupling the proposed localisation approach with an iterative segmentation method might lead to a fully automated and precise approximation of the FLL boundaries, allowing the radiologists to use an automatic delineation of the FLL for its assessment.

References

1. Ambai, M., Yoshida, Y.: CARD: Compact and Real-time Descriptors. In: IEEE International Conference on Computer Vision. pp. 97–104 (2011)
2. 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 BMVA 2012(7), 1–14 (2012)
3. Bakas, S., Hoppe, A., Chatzimichail, K., Galariotis, V., Hunter, G., Makris, D.: Focal Liver Lesion Tracking in CEUS for Characterisation Based on Dynamic Behaviour. In: Bebis, G., et al (eds.) Advances in Visual Computing. LNCS, vol. 7431, pp. 32–41. Springer Berlin Heidelberg (2012)
4. Bakas, S., Hunter, G., Thiebaud, C., Makris, D.: Spot the Best Frame: Towards Intelligent Automated Selection of the Optimal Frame for Initialisation of Focal Liver Lesion Candidates in Contrast-Enhanced Ultrasound Video Sequences. In: 9th Int. Conf. on Intelligent Environments. pp. 196–203. IEEE Press (2013)
5. Bakas, S., Sidhu, P.S., Sellars, M.E., Hunter, G.J.A., Makris, D., Chatzimichail, K.: Non-invasive Offline Characterisation of Contrast-Enhanced Ultrasound Evalu-

ations of Focal Liver Lesions: Dynamic Assessment Using a New Tracking Method. In: 20th European Congress of Radiology (2014)

6. Claudon, M., Dietrich, C.F., Choi, B.I., Cosgrove, D.O., Kudo, M., Nolsoe, C.P., *et al.*: Guidelines and Good Clinical Practice Recommendations for CEUS in the Liver - Update 2012: A WFUMB-EFSUMB Initiative in Cooperation with Representatives of AFSUMB, AIUM, ASUM, FLAUS and ICUS. Ultrasound Med Biol 39(2), 187–210 (2013)

7. Crum, W.R.: Spectral Clustering and Label Fusion for 3D Tissue Classification: Sensitivity and Consistency Analysis. Annals of the BMVA 2009(6), 1–12 (2009)

8. Figueiredo, M.A.T., Jain, A.K.: Unsupervised Learning on Finite Mixture Models. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(3), 381–396 (2002)

9. Harvey, C.J., Blomley, M.J.K., Eckersley, R.J., Cosgrove, D.O.: Developments in Ultrasound Contrast Media. Eur Radiol II, 675–689 (2001)

10. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)

11. Otsu, N.: A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man and Cybernetics 9(1), 62–66 (1979)

12. Rognin, N., Campos, R., Thiran, J.P., Messager, T., Broillet, A., Frinking, P., *et al.*: A New Approach For Automatic Motion Compensation For Improved Estimation of Perfusion Quantification Parameters in Ultrasound Imaging. In: 8th French Conference on Acoustics. pp. 61–65 (2006)

13. Schneider, M.: Charasteristics of SonoVue. Echocardiography 16(7), 743–746 (1999)

14. Shekhar, R., Zagrodsky, V.: Mutual Information-Based Rigid and Nonrigid Registration of Ultrasound Volumes. IEEE Transactions on Medical Imaging 21(1), 9–22 (2002)

15. Song, Z., Tustison, N., Avants, B., Gee, J.: Adaptive Graph Cuts with Tissue Priors for Brain MRI Segmentation. In: 3rd International Symposium on Biomedical Imaging: Nano to Macro. pp. 762–765. IEEE Press (2006)

16. Strobel, D., Seitz, K., Blank, W., Schuler, A., Dietrich, C.F., vonHerbay, A., *et al.*: Tumor-Specific Vascularization Pattern of Liver Metastasis, Hepatocellular Carcinoma, Hemangioma and Focal Nodular Hyperplasia in the Differential Diagnosis of 1349 Liver Lesions in Contrast-Enhanced Ultrasound (CEUS). Ultraschall Med 30(4), 376–382 (2009)

17. Ta, C.N., Kono, Y., Barback, C.V., Mattrey, R.F., Kummel, A.C.: Automating Tumor Classification With Pixel-by-Pixel Contrast-Enhanced Ultrasound Perfusion Kinetics. J Vac Sci Technol B Nanotechnol Microelectron 30(2), 02C103 (2012)

18. Westwood, M.E., Joore, M.A., Grutters, J.P.C., Redekop, W.K., Armstrong, N., Lee, K., *et al.*: Contrast-Enhanced Ultrasound Using SonoVue®(Sulphur Hexafluoride Microbubbles) Compared With Contrast-Enhanced Computed Tomography and Contrast-Enhanced Magnetic Resonance Imaging for the Characterisation of Focal Liver Lesions and Detection of Liver Metastases: a Systematic Review and Cost-Effectiveness Analysis. Health Technol Assess 17(16), 1–243 (2013)

19. Wilson, S.R., Burns, P.N.: Microbubble-Enhanced US in Body Imaging: What Role? Radiology 257(1), 24–39 (2010)

20. Zhou, J., Huang, W., Xiong, W., Chen, W., Venkatesh, S.K., Tian, Q.: Delineation of Liver Tumors from CT Scans Using Spectral Clustering with Out-of-Sample Extension and Multi-Windowing. In: Yoshida, H., Hawkes, D., Vannier, M. (eds.) Abdominal Imaging, Computational and Clinical Applications. LNCS, vol. 7601, pp. 246–254. Springer Berlin Heidelberg (2012)