Abstract—Parameters extracted from time activity curves in contrast enhanced ultrasound images play an important role in independent or computer aided diagnosis of hepatic focal lesions. Due to noise and errors induced by movement of ultrasound probe and breathing of patients, reproducible extraction time activity curve parameters is challenging. In this paper we propose a new solution that combines filtering and tracking, in the theoretical framework of robust estimation and the mean shift. To cope with the gradual appearance changes of the tracked lesions, we propose a new dynamic scale selection method.

I. INTRODUCTION

Ultrasonography is an inexpensive and non-invasive medical imaging technique, which has been used successfully to assess thoracic and abdominal organ diseases. Contrast Enhanced Ultrasound (CEUS) imagery is a newer technique, using a contrast enhancer agent consisting of micro-bubbles [1-4]. Micro-bubbles have diameters of about 10 micrometers and act as amplifier agents of the blood backscattering. The agent has to be injected into the blood stream of the patient just before examination. Micro-bubbles are safe on patients, even with renal or liver failure, and are eventually excreted through respiration and breakdown in the liver. This process takes about 3 to 5 minutes. Micro-bubbles have a high degree of echogenicity and make possible the evaluation of blood concentration in tissues. Blood perfusion kinetic changes can be useful for the detection of possibly malignant tissue. The diagnosis is primarily based on the evaluation of several parameters of perfusion dynamics characterized by parameters extracted from the time intensity curves (TIC). Such curves are computed in a region of interest level and subsequently used in evaluation of hepatic liver lesions [5], renal disorders [6-8] or heart diseases [9].

CEUS examinations are performed by the sonographer operator free hand. In practice, although the operator usually tries to keep the probe in a steady position, there is a significant amount of motion in the acquired image sequences, also due to patient breathing motion. Sometimes motion is made on purpose by the sonographer operator, in order to avoid early bubble destruction caused by the ultrasound waves.

The recorded dataset consist of two image sequences: the contrast enhanced (CE) image and the basic ultrasound image (the so called B mode image). The images frames are obtained alternately, at a rate of about 10 frames per second and separated by an acoustic filtering process. The B mode image is showing the anatomy more clearly, while the CE image is showing the functional information. Both images contain a high level of speckle noise, in comparison with images acquired by computer tomography (CT) or magnetic resonance imaging (MRI). Given the small size of the bubbles and the finite resolution of the scanners, enhanced images contain a specific type of noise, with a switch on-off character which can be explained as effects of under-sampling. This noise, along with motion poses important challenges on the CEUS image processing task.

In this paper, we propose a new method for TIC extraction for focal liver lesion (FLL) analysis in CEUS images, by a joint tracking and filtering process, funded on robust statistics and the mean shift algorithm. The paper is organized as follows. In the second section, we review previous approaches in image sequence filtering, from the perspective of medical imaging. In the third and fourth section, we present and motivate our proposed approach. The results of our experiments are included in the fourth section, while the last section is devoted to conclusions.

II. METHODS

A theoretical model of the TIC and a graphical definition of several parameters of the curve needed for objective quantification of the blood perfusion kinetics, are given in Fig. 1.
Parameters extracted from TICs play an important role in the diagnosis of CEUS studies. Unfortunately most of them are extremely sensitive to noise.

General image denoising techniques can be useful in order to obtain smoother TICs. To avoid contamination of FLL data from neighbouring pixels, edge preserving smoothing methods, such as the bilateral filter [10], the mean shift filter [11] or the non-local means filter [12] can be used.

One powerful tool for generating smooth TICs, suitable for dynamic parameter extraction can be obtained by regression techniques. A widely used technique, often included in the software packages developed by scanner manufacturers, is based on fitting the data to a regular curve. Exponential fitting of the arterial phase was used in [9],[5]. For global TIC modelling, with both arterial and venous (portal) phases, in [13], a log normal distribution model is used. While, by using robust fitting, this approach may be working effectively and naturally being able to tolerate missing samples, in some real-life applications, the curve generated by this approach is not accurate enough to enable unbiased parameter extraction [14].

A major source of error in TIC extraction is caused by motion. Unlike the speckle noise characteristic to US images, errors induced by motion are not random and therefore violate the usual assumptions used in optimizing the image filtering methods. Instead of limiting the processing work to finding intelligent curve fitting solutions, it is wiser to prevent motion errors to get into the dataset before the TIC was generated. The most straightforward methods of coping with motion errors are image segmentation and image registration for motion compensation.

Interactive image segmentation is the most accurate solution to CEUS image segmentation, but nevertheless too tedious to be routinely done by physicians. To alleviate the segmentation problem, in [14], a user-friendly high performance software for image segmentation and analysis is developed. The authors define spatio-temporal volumes of 5x5x3 voxels, wherein they compute enhancement values. Voxels with similar enhancement profiles are subsequently grouped in an interactive segmentation procedure. Since the enhancement images contain very high levels of noise and the neighbourhood size used is small, the TICs are generated by means of robust averaging methods. The first and the third quartile averages define lower and upper TIC envelopes, while the percentage of pixels with enhancement level higher than a threshold, k, generated an enhancement TIC profile which proved to be the most reliable measure. The optimal neighbourhood size and the optimum threshold remained to be set interactively by the user.

Automatic segmentation, that would be a valuable help in terms of reducing the burden of interactive work for the physicians, is problematic with CEUS images, due to the high noise level and the weak contrast of lesions in the early wash-in and in the late wash-out stages of the image sequence. B mode images could be used instead for lesion localization, but this works only for FLLs which are visible in the base images. A fully automatic image segmentation method to detect renal anechoic lesions in CEUS images was proposed in [15]. A guided filter [16] is used as a pre-processing step. No TIC analysis is performed, since only the lesions are segmented, not the whole kidney region.

Image registration is another approach for coping with motion effects. Many rigid and deformable image registration techniques have been developed for biomedical imaging purposes, including some devoted to US images [17][18]. A work optimised for the new class of CEUS images is reported by [19]. Registration techniques are also adopted in [20], where a Markov Random Field model is used to optimize image registration for intestinal wall analysis. The improvement of image registration is measured by the reduction of the standard deviation of the registered regions, with about 8%.

### III. MOTIVATION AND THEORETICAL BACKGROUND OF THE MEAN SHIFT ALGORITHM

Reliable extraction of TICs of FLLs from CEUS data requires processing methods that are able to cope effectively with the uncertainty generated by heavy noise and target motion. Ideally, the intensity of the enhancement image should be evaluated only within the lesion region. The region could be specified through interactive or automatic segmentation. However, none of these is error free. Even with a good segmentation, the speckle noise generates many data samples that deviate far from the true value within the region of interest. Such samples, called outliers, when asymmetrically distributed, may produce severe bias of simple estimators like the arithmetic mean, or the weighted mean. Segmentation errors, improper motion estimation and target deformation are typical sources of asymmetrically distributed outliers.

Methods being able to cope effectively with contamination of the data with outliers are called robust. All robust methods have in common the characteristic that outliers have limited influence of the estimates produced, no matter how much they deviate from the true value, whereas inliers have a much heavier influence. Among the wide range of robust image processing methods designed within the community of computer vision, applications of the mean shift algorithm [21] in image filtering [11], segmentation [11],[22] and tracking [23][24] received considerable attention in the last decade.

The mean shift algorithm is an efficient way of finding probability density modes within a discrete set of multivariate data, without having to estimate the whole probability density function. Starting from any location of the feature space, for example starting from a particular sample point, the algorithm computes a weighted mean of the samples around the starting point and moves to the weighted mean found. The procedure is repeated from the newly found point and iterated until convergence. It can be shown that the step, which is the mean shift vector, is in the direction of the density gradient and the density of the newly found point is higher or equal to the gradient at the initial point.

Let \( \mathbf{x} \) be vector data samples from a data set of size \( n \), \( w() \) a continuous, 1D function with a smoothing role, \( \mathbf{H} \) a bandwidth matrix with the role of scaling the vector components and \( d(x,H) \) a scaled measure of distance in the feature space. In monochromatic images, data samples are vectors with three components: two of them specifying the space coordinates and one of them the image value, or
the range information. For a pixel with space coordinates \((l,c)\) and intensity \(f(l,c)\), the vector is \(x=[l,c,f(l,c)]^T\). Most often than not, for simplicity, the bandwidth matrix used has a diagonal form.

The constant bandwidth mean shift algorithm can be summarized as follows.

1. Set the starting point, \(y_1=x\).
2. Set \(j=1\).
3. Repeat equation 1,
   \[
   y_{j+1} = \frac{\sum_{i=1}^{n} x_i w(y_j - x_i, H)}{\sum_{i=1}^{n} w(y_j - x_i, H)}, \quad j = 1, 2, \ldots
   \]
   until convergence, i.e. \(\|y_{j+1} - y_j\| < \epsilon\), a small threshold.
4. Set the convergence point, \(y_{j+1}\) as the final result.

Since the density gradient is zero at any density mode, the mean shift algorithm is climbing the density space to a local mode and stops there. As the algorithm proceeds to higher density regions of the feature space, the mean shift vector magnitude is gradually reduced, a favourable property, resulting in monotonic convergence to a density mode.

The mean shift algorithm can be derived either in the framework of probability density estimation or in the framework of M estimators [25]. In the first case \(w()\) is the kernel profile, which is the derivative of the smoothing kernel used for density estimation. In the second case, \(w()\) is the derivative of the loss function. The main equation of the mean shift algorithm computes a weighted average of the data samples. The denominator of equation has the role to normalize the sum of weights to 1. Weights for each sample depend on the scaled difference between the current estimate and the sample. The weighting function is designed such that, the higher the difference, the less the weight of a sample. This feature ensures that outlier samples have a small weight, therefore a small influence on the estimate.

In image filtering, the mean shift algorithm starts with the first estimate as the current pixel data point. The result is the range component of the convergence point of the algorithm. Note that, during iterations, as the estimate vector changes, the space component of the estimate may change, thus moving away from the initial location of the pixel, toward the location with (local) maximum density of similar pixels. This property can actually be used to track a homogeneous target on a variable background. If the target is heterogeneous, the mean shift tracker can be designed to follow the target by finding the location minimizing the dissimilarity between the local and the target histograms [26]. Since, in this work, the target is a hyperechoic focal liver lesion, which is fairly homogeneous, we work directly on the intensity image data.

IV. USING THE MEAN SHIFT ALGORITHM FOR TRACKING FLLS AND EXTRACTING TICS IN CEUS

One critical problem with applying the mean shift algorithm to track FLLs is generated by the fact that, unlike in previous applications the target is not only moving, but also constantly changing its intensity.

Moreover, the target is not visible at all in the first few tens of frames. To overcome the second problem, we initialize the tracker by selecting a convenient frame, where the target is clearly visible and interactively setting the ROI, as illustrated in Fig. 2. The first and more difficult problem is approached by using a dynamic scale parameter for the range component of the data. As the target size remains constant, the scale parameter for the space information is kept constant.

We propose a separable weighting function of the form:

\[
\begin{align*}
    w_l(c,l,f) &= w(c)w(l)w_f(f), \\
    w(c) &= \begin{cases} 
        1 & |c/h_{sc}| < 1 \\
        0 & \text{otherwise}
    \end{cases}, \\
    w(l) &= \begin{cases} 
        1 & |l/h_{sd}| < 1 \\
        0 & \text{otherwise}
    \end{cases}, \\
    w_f(f) &= \begin{cases} 
        1 & |f/h_r(k)| < 1 \\
        0 & \text{otherwise}
    \end{cases}
\end{align*}
\]

In (2), \(h_{sc}, h_{sd}\) are space scale parameters. The dynamic scale mean shift algorithm proposed is summarized below. Let \(f(c, l, k)\) be a pixel intensity at space coordinates \((c,l)\) and frame \(k\). Let \(K\) the number of frames in the study.

1. Set the initial mean shift estimate of the starting frame, \(k\), as:
   \[
   \left[ c_1^k, l_1^k, f_1^k \right]^T = \begin{bmatrix} \text{median}(c \in \text{ROI}) \\ \text{median}(l \in \text{ROI}) \\ \text{median}(f(c,l,k) \mid (c,l) \in \text{ROI}) \end{bmatrix}
   \]

Define the neighbourhood of any point, \((c,l)\), \(N(c,l)\) as a rectangular region of the ROI’s size, with the centre in \((c,l)\).

2. for \(i = k \) to \(K-1\) do
   2.1 Set \(j = 1\).
   2.2 Repeat until convergence:
   \[
   \left[ c_{j+1}^k, l_{j+1}^k, f_{j+1}^k \right]^T = \begin{bmatrix} 
   \text{median}(c \in \text{ROI}) \\ \text{median}(l \in \text{ROI}) \\ \text{median}(f(c,l,k) \mid (c,l) \in \text{ROI}) \end{bmatrix}
   \]
where \(j=1,2,\ldots\)

2.3 Set
   \[
   \left[ c_{conv}^k, l_{conv}^k, f_{conv}^k \right]^T = \left[ c_{j+1}^k, l_{j+1}^k, f_{j+1}^k \right]^T
   \]
3. for $i = k-1$ to 0 do steps 2.1 to 2.3. 

The generated TIC is then:

$$f_{\text{conv}}^k, \quad k = 1, 2, ..., K - 1$$

(6)

While the particular shape of the weighting function is of secondary importance, the scale parameters set the degree of density smoothing and, therefore, have the strongest influence on the results of the algorithm. The scale establishes how fast the influence of a sample decreases as a sample is farther away from the current estimate in space or in range.

A natural choice for the space scale parameters is

$$h_x = C / 2$$
$$h_y = L / 2$$

(7)

where $C$ and $L$ are column and line numbers of the ROI. With this choice, the space weights take on the value 1 at the space coordinates of the current mean shift estimate and decrease to zero at locations reaching the borders of the ROI, centered on the space coordinates of the current mean shift estimate. As pixels far from the current spatial centre of the lesion have small influence both on the next spatial centre and on the next estimate of the frame intensity around that centre, exact extraction of the FLL region is less critical for the accuracy of the generated TIC. This is why, we use a rectangular space window.

Given the high dynamic of the target intensity during the study, selection of the range scale parameter is more critical. The scale sets the selectivity of the tracker for the intensity information. More specifically, with the kernel adopted in this work, samples with absolute intensity deviation from the current estimate higher than the range scale, $h_r$, have zero influence on the next estimate. A small scale rejects better the noise. However, the scale should not reject samples with a moderate amount of noise, since these are valuable in the accurate estimation of the lesion intensity. The range scale also establishes the maximum of the interframe intensity variation that the tracker is able to follow, so, again, a too selective scale is not desirable. Yet another concern in scale selection is the gradual loss of contrast as the tracker goes back from the selected frame toward the initial frame. In these frames, a high range scale allows too many background pixels to contribute to the estimate. As a result of the low contrast and the superior number of background pixels around the lesion, the tracker may be easily attracted into the background and target lost. In such cases, the first part of the generated TIC drops to zero too fast and this distortion affects drastically many of the parameters extracted from the TIC, namely the TOA, the RT, and the TTP.

To meet all the needs mentioned above, we propose a dynamic range scale selection of the form:

$$h_r = \left( s_0 + \frac{k}{K} s_1 \right) f_k$$

(8)

In the last equation, $s_0$ and $s_1$ are two constants, setting the minimum and maximum fraction of the estimate used to compute the range scale. This way, the range scale automatically adapts to the intensity level and the noise level (which grows with the intensity). The time dependency induced by the frame index, $k$, makes the range scale to be a changing fraction of the lesion intensity. This feature allows the design of a more selective tracker when it goes back in time from the starting frame toward the initial frames, where the lesion is barely detectable even by the physicians.

V. EXPERIMENTS

The mean and median intensities in the ROI can be used fairly well to extract the curve, in the absence of movement. However their accuracy is severely impaired when there is significant movement of the focal liver lesion with respect to the probe. We use the TICs generated by mean and median extraction in this work just in order to see the gain obtained by the proposed algorithm. In Fig. 3 and Fig. 4, a TIC generated with a fixed ROI with the mean respectively the median estimate is illustrated, for an image sequence with significant probe movement. The effect of movement is well noticeable on the TIC. The main movement occurs between frames 310 to 390, as illustrated in Fig. 5. The TIC generated for the same sequence with the mean shift tracker is shown in Fig. 6. The curve is much smoother and more suitable to further analysis and parameter extraction. To check how the mean shift tracker is able to compensate the movement, in Fig. 7, we marked the tracker space coordinate estimates by red dots. Clearly, the horizontal trajectory curve demonstrates that the lesion movement is correctly tracked.

Figure 3. Generated TIC using the mean

Figure 4. Generated TIC from CEUS sequence using median estimate
VI. CONCLUSIONS

In this paper, we designed a mean shift based tracker which is able to track FLLs in CEUS image sequences with significant probe movement, for accurate TIC extraction. To overcome the challenge posed by the fact that the lesion is permanently changing its intensity between consecutive frames, we proposed a dynamic range scale parameter selection method. Further curve filtering by model fitting may be used prior to parameter estimation. Parameters extracted from TICs of the FLLs are important in the computer aided differential diagnosis of focal liver lesions.

ACKNOWLEDGMENT

“This work was partially supported by the strategic grant POSDRU 107/1.5/S/77265 (2010) of the Ministry of Labour, Family and Social Protection, Romania, co-financed by the European Social Fund – Investing in people.”

REFERENCES

[3] Ioan Sporea et al., Contrast enhanced ultrasound for the characterisation of focal liver lesions, 2011, Medical Ultrasonography 13(1):38-44


