Classification of kidney and liver tissue using ultrasound backscatter data

Fereshteh Aalamifar, Hassan Rivaz, Juan J. Cerrolaza, James Jago, Nabile Safdar, et al.

Event: SPIE Medical Imaging, 2015, Orlando, Florida, United States
Classification of Kidney and Liver Tissue Using Ultrasound Backscatter Data

Fereshteh Aalamifar¹,², Hassan Rivaz³, Juan J. Cerrolaza², James Jago⁴, Nabile Safdar², Emad M. Boctor¹, Marius George Linguraru²,⁵

¹Johns Hopkins University, Baltimore, MD, USA
²Sheikh Zayed Institute for Pediatric Surgical Innovation, Children’s National Health System, Washington DC, USA
³Concordia University, Montreal, QC, Canada
⁴Philips Healthcare, Bothell, WA, USA
⁵George Washington University Medical School, Washington DC, USA

ABSTRACT
Ultrasound (US) tissue characterization provides valuable information for the initialization of automatic segmentation algorithms, and can further provide complementary information for diagnosis of pathologies. US tissue characterization is challenging due to the presence of various types of image artifacts and dependence on the sonographer’s skills. One way of overcoming this challenge is by characterizing images based on the distribution of the backscatter data derived from the interaction between US waves and tissue. The goal of this work is to classify liver versus kidney tissue in 3D volumetric US data using the distribution of backscatter US data recovered from end-user displayed B-mode image available in clinical systems. To this end, we first propose the computation of a large set of features based on the homodyned-K distribution of the speckle as well as the correlation coefficients between small patches in 3D images. We then utilize the random forests framework to select the most important features for classification. Experiments on in-vivo 3D US data from nine pediatric patients with hydronephrosis showed an average accuracy of 94% for the classification of liver and kidney tissues showing a good potential of this work to assist in the classification and segmentation of abdominal soft tissue.

Keywords: Tissue characterization, ultrasound imaging, machine learning, computer-aided diagnosis, liver, kidney.

INTRODUCTION
Ultrasound (US) is an easy-to-use, inexpensive, real-time and safe imaging modality that is routinely used for diagnosis and treatment planning. However, the quality of US data is highly dependent on the talent and expertise of the sonographer, and the interpretation of US images is difficult, thus usually qualitative and subjective. Moreover, most clinical practices use 2D US images, which present only cross-sections of anatomy and increase the variability in the interpretation of these data. Volumetric 3D US can improve the quality of image acquisition and decrease the amount of subjectivity and variability of the image interpretation [1]. Also, by adding automatic classification and segmentation of organs and tissue in 3D US images, we can enable volumetric and quantitative analysis [2, 3]. In this paper, we propose the computation of a large set of US features for the characterization of normal soft tissue in 3D US images based on the distribution and correlation of speckles in the backscatter data. We analyze the relevance of these features for the classification of liver and kidney tissue and validate the method on a dataset acquired for a pediatric application.
US soft tissue characterization can be done based on the distribution of US speckle data. The envelope detected US data is known to have homodyned-K distribution [4, 5]. The parameters of this distribution depend on the tissue type and therefore can be used as features for tissue classification [6]. On one hand, analyzing the parameters of the homodyned-K distribution requires a large number of samples to make the statistical inference reliable. On the other hand, it is desirable to perform tissue classification based on a small number of samples to create high-resolution parametric US images. Hence, it is preferable to exploit more features. Previously, speckle correlation coefficients between selected patches of image frames have been used for sensorless tracking of 2D US images [7-9]. Furthermore, in our previous work [7], we found that the correlation coefficients depend on the tissue type thus making the tracking information unreliable. In this work, we do the opposite and benefit from the correlation coefficients to infer tissue properties. The advantage of such feature computation is that it is possible to compute features in all directions of the 3D image or by comparing consecutive or non-consecutive frames yielding a comprehensive set of features.

Even though the proposed features have the potential to differentiate between any tissue types, in this paper, we focus on a specific clinical application: kidney and liver tissue differentiation in 3D US abdominal images from pediatric patients with hydronephrosis. Hydronephrosis is a common disease in pediatric patients, affecting 2-2.5% of children [10]. US imaging is useful in the early diagnosis of hydronephrotic patients, but yet is limited by its subjective and qualitative assessment. Thus, US tissue characterization can help with the initialization of automatic segmentation algorithms [11] and can also be beneficial in providing further information for the diagnosis of hydronephrosis.

**MATERIALS AND METHODS**

In this section, different parts of the proposed framework for tissue classification are described. Firstly, the US image is decompressed to estimate the envelope detected backscatter data. Then, the homodyned-K and correlation coefficient features are extracted. A random forests classifier is used to both select relevant features for abdominal soft tissue characterization and perform cross validation. Fig. 1 shows an overview of the training and validation process.

![Flowchart](image)

**Figure 1.** An overview of the process for tissue characterization; two groups of features are extracted from the decompressed image (homodyned-K parameters and correlation coefficients); then random forests classifier is used to determine relevant features and cross validation is performed to evaluate the effectiveness of the method.

**US Image Decompression**

In most clinically acquired US images, logarithmic compression has been performed on the original US data to reduce the dynamic range of the envelope data, which causes the speckle distribution to be no longer homodyned-K [4]. To recover the original envelope data, we used the decompression method proposed by Seabra et al. [12] to estimate the envelope-detected data (decompression parameter). Assuming the decompression parameters are constant inside a small region, we used a sliding-window-based approach to estimate the parameters throughout the image. We then averaged the locally computed decomposition parameters to yield the estimated decompression parameter for the entire US image. Considering boxes of size $n_x \times n_y \times n_z$, ($n_x = n_y = n_z = 15$) the final decompression parameter, $D$, was calculated using the analytical formula

$$D = \frac{1}{N_x N_y N_z} \sum_{i,j,k}^{N_x N_y N_z} \frac{\sqrt{24} \sigma_{i,j,k}}{\pi}, \quad (1)$$

where $N_x$, $N_y$, $N_z$ are the numbers of boxes along the three axes of the image, respectively and $\sigma_{i,j,k}$ is the standard deviation of voxel values in each box. After the estimation of $D$, we estimated the envelope data as:
\[ E_{x,y,z} = e^{c_{x,y,z} t_D}, \]  

(2)

where \( c_{x,y,z} \) is the voxel value of the compressed image at pose \((x, y, z)\). We performed the decompression before feature extraction.

**Homodyned-K Distribution Features**

Speckles of envelope US data were shown to have homodyned-K distribution [4, 13]. The probability function of the homodyned-K distribution has three parameters: \( \alpha \), \( \alpha \), and \( k \) (mean signal energy, effective number of scatterers per resolution cell, and ratio of coherent to diffuse scattering, respectively). \( \alpha \) and \( k \) (which we call H-K features) may be used for tissue classification [6, 14, 15]. We used the algorithm proposed in [6] to estimate these two parameters. This method estimates the homodyned-K parameters based on crossing level curves of the signal-to-noise-ratio (R), skewness (S), and kurtosis (K), (called the RSK algorithm) deduced from the fractional order moments of the envelope-detected data.

In [6], it is shown that the three statistical parameters, RSK, are a function of only \( \alpha \) and \( k \) of the underlying homodyned-K distribution, and the moment order, \( v \). The moment order can be chosen arbitrarily, however, the authors in [6] recommend using the moments of order 0.72 and 0.88 as the “optimal” ones. Hence, RSK algorithm gets a patch of envelope detected samples (a patch of decompressed voxel values) as input and estimates the three statistical parameters; then it finds the level curves of these three parameters in the Cartesian space of \( \alpha \) and \( k \). The nearest point to all the three level curves is the estimated \( \alpha \) and \( k \), namely the output of the RSK algorithm.

If the RSK algorithm fails to find \( k \) or \( \alpha \), in a reasonable parameter range, we set them to -1 as an indication of number of failures. In our implementation, we used patch sizes of \( n \times m \times p \) voxels with \( n=m=p=20 \). Choosing the right patch size is a trade-off between the accurate estimation of H-K features and a good resolution of the parametric image. In this paper, pilot experiments were performed to define the appropriate size for our method, though further investigation may be required in order to optimize the patch size for specific applications.

**Speckle Correlation Coefficient Features**

Correlation coefficients between speckles on US image frames were previously used for sensorless tracking of 2D US images [7-9]. In speckle tracking applications, the correlation coefficients were usually computed for patches of fully developed speckles because these coefficients change for different tissue types [7]. We use this property of the US speckle to create additional features for tissue characterization. We divided the 3D US image into patches of size \( n \times m \times p \) \((n=m=p=20)\). In each patch, we considered two 2D slices \( v_1 \) and \( v_2 \), and computed the Pearson correlation coefficient \( \rho \),

\[
\rho = \frac{\sum_{i=1}^{N} (v_{1i} - \bar{v}_1)(v_{2i} - \bar{v}_2)}{\sqrt{\sum_{i=1}^{N} (v_{1i} - \bar{v}_1)^2 \sum_{i=1}^{N} (v_{2i} - \bar{v}_2)^2}}
\]

(3)

where \( \bar{v}_1 \) and \( \bar{v}_2 \) are the mean voxel values, and \( N \) is the number of voxels in each slice. We selected \( v_1 \) and \( v_2 \) in three directions as shown in Fig. 2(a). Furthermore, we define \( v_1 \) and \( v_2 \) in two different ways: (1) by fixing \( v_1 \) to the first slice in the US volume and sliding \( v_2 \), and (2) taking \( v_1 \) and \( v_2 \) consecutively and sliding them together. We denote these features as \( \rho_{sl} \) and \( \rho_{con} \) respectively. Therefore, given a patch of size \( n \times m \times p \) there are \( n+m+p-3 \) features of each class, i.e., \( \rho_{sl} \) and \( \rho_{con} \). Including the H-K features, the total number of tissue characterizing feature is \( 2 \times (n+m+p)-4 \) for each image patch.
Random Forests

To evaluate the potential of the new features to discriminate between liver and kidney tissue, we use a random forests-based classifier [16]. Random forests is a popular ensemble learning algorithm [17, 18]. The advantages of using the random forests framework include the following: 1) it can handle a large set of features without overfitting; 2) it is not resource intensive in terms of memory and time in our tests; and 3) it computes feature relevance (importance) for the classification and unbiased error as it constructs the decision trees without the need for a separate test set.

During the training process, random forests algorithm constructs many decision trees. When constructing a tree, at each splitting node, the algorithm chooses a subset of features randomly (instead of using all the features). During the test process, random forests, puts the test case down all the decision trees and collects the votes from the trees. The class that has the most number of votes will be assigned to the test case. Hence, random forests requires two main parameters: 1. the number of grown trees, 2. the cardinality of the features subset (the number of features to select at each splitting node). The first parameter usually only determines the required time and memory for the algorithm to run. The second parameter determines the correlation between the trees and also the strength of each tree. As the cardinality of the features’ subset increases, the correlation between the trees increases, and the strength of the trees increases too. However, to have a stronger classifier, one would want to minimize the correlation between the trees and maximize the strength of each individual tree. Hence the second parameter should be chosen carefully. There are other parameters in random forests algorithm that can be set for each application. In our case, the samples were unbalanced, i.e., there was a much larger number of kidney patches available than of liver patches. To address this issue, a parameter called cut-off can be used. This parameter determines the portion of votes for each class that needs to be achieved in order to win. We chose 500 as the number of grown trees, 10 as the cardinality of the features subset, and [0.3, 0.7] as the cut-off parameters for [liver, kidney].

Feature Selection

As mentioned before, the random forests algorithm provides feature relevance while constructing the decision trees. We ranked features based on the mean decrease in Gini index [19, 20] provided by the random forests algorithm. To determine the number of features to select for optimal classification, we used this ranking together with the area under the receiver operating characteristic curve (AUROC) [21]. We selected the least number of features that maximized AUROC.

EXPERIMENTS AND RESULTS

The clinical 3D US images were collected using a Philips iU22x Matrix system (Amsterdam, Netherlands) with X6-1 Matrix array probe. The acquisition and processing of data were approved by the institutional review board at Children’s National Health System. Right kidney images, as shown in Fig. 2(b), were taken from nine pediatric patients who were diagnosed with hydronephrosis. Images were segmented manually by creating a 3D mask. The entire kidney was segmented by a radiologist and part of the liver, only partially visible in these images, was segmented by a research fellow under the supervision of a radiologist. The liver segmentation was not performed to accurately
define the boundaries; the purpose was rather to identify tissue areas that are inside the liver in order to train and test our proposed classification algorithm.

Figure 3: Classification features: (a) H-K features extracted from all samples (10^5-10^6 voxels) of kidney and liver masks in 9 right kidney 3D US images. (b) H-K features extracted from 8000 voxels resembling a patch size of 20x20x20 voxels. (c) Two \( \rho_{x2} \) features extracted from all patches of kidney and liver masks along the X axis.

Experimenting with Features

The above explained decomposition method was applied to the whole US image volume to estimate the envelope detected data. The RSK method was used to estimate the homodyned-K parameters for all the samples from the kidney and liver masks. Fig. 3(a) shows the extracted parameters for the 9 hydronephrotic patients using all the voxels inside the kidney or liver regions of each image. It can be seen that the two H-K features are clearly differentiating between the liver and the kidney. Moreover, the range of values of \( k \) is markedly different for both tissues (a threshold of \( k=0.4 \), allows to divide the samples into two classes, kidney and liver, with 100% accuracy). However, in Fig. 3(a), all voxels (around 10^5-10^6 voxels per kidney or liver per case) were used to estimate the H-K parameters demonstrating the potentiality of these parameters for normal tissue classification. In a more typical scenario, it is desirable to create parametric images by dividing the volume into smaller patches. Fig. 3(b) shows the extracted parameters for 8000 voxels resembling a patch size of 20x20x20 voxels. It can be seen that, as the number of voxels in each patch is decreased, it becomes more difficult to do the classification (83% classification accuracy by thresholding at \( k=0.4 \)).

The second set of features was extracted through speckle correlation coefficients. To analyze the potential of these features, the 2D slices along the Y-Z axis were compared and analyzed (when the slice is comprised of voxels along Y-Z axes for a constant X, the features are called to be along X axis, see also Fig 2(a)). Fig. 3(c) shows the correlation coefficients of the 6th and 3rd slices when compared to the first slice in the image, as an example. Even though the correlation coefficients are overlapping for the two tissue classes, they bring additional differentiating capability, as exemplified in Fig. 3(c) which was created using only two correlation coefficient features.

Classification and Validation

The feature extraction process was performed as follows: 1) the largest cube inside each of the liver or kidney masks was divided into patches of size 20x20x20 voxels; 2) for each patch, the H-K parameters were calculated; 3) then, the correlation coefficient features were extracted. For \( \rho_{x2} \) features, the first patch correlation with itself was also included, yielding a total of 119 combined features for each patch; this inclusion allows us to see how random forests produce features important for an obvious non-important feature.

We used the random forests implementation in [20] to do the classification. As mentioned before, since the number of liver patches in the current segmentation was smaller than that of the kidney patches, the random forests classification was done with a cut-off parameter set to 0.3 for liver, i.e. liver wins if it has only 0.3 of the votes from the trees. Then, the features’ relevance in terms of mean decrease in the Gini index was evaluated by using all the available patches (as shown in Fig. 4(b)). The AUROC was evaluated to select the relevant features for tissue classification, as shown in Fig. 4(a). In Fig. 4(a), the maximum AUROC value was 0.98 (considering two decimal digits). We selected the minimum number of features yielding AUROC=0.98 (20 features). In order of decreasing importance, the most relevant features were \( k \), 16 of the \( \rho_{x2} \) features, and finally 3 features from the \( \rho_{x1} \) set.
Figure 4: Feature selection. (a) AUROC values when using various number of features. (The features are ranked based on their relevance shown in (b)). (b) Feature relevance in terms of mean decrease in Gini index for patches of size 20x20x20 extracted from kidney and liver masks; Larger values depict higher importance. Red bins show the selected 20 features.

Cross Validation

We employed the cross validation [17] to evaluate the classification. Our image database was collected mainly to observe the kidney and hence, only part of the liver was visible and also partially segmented. In addition, because we considered cubic patches, consisting of 20x20x20 voxels in all the three axes of the image, some images had very limited number of liver patches and even none. Since the number of available liver patches was very limited, we performed cross validation after dividing the images into groups, trying to evenly distribute the number of kidney and liver patches among the groups. At each time, one group was set aside for test and the rest were used for training. We did the grouping based on the number of patches belonging to each class. Table 1 shows the number of patches for each of the 9 images.

Table 1. Number of patches in each image.

<table>
<thead>
<tr>
<th>Image #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kidney</td>
<td>58</td>
<td>26</td>
<td>29</td>
<td>25</td>
<td>23</td>
<td>42</td>
<td>125</td>
<td>67</td>
<td>92</td>
<td>487</td>
</tr>
<tr>
<td>Liver</td>
<td>2</td>
<td>1</td>
<td>12</td>
<td>0</td>
<td>6</td>
<td>16</td>
<td>0</td>
<td>12</td>
<td>15</td>
<td>64</td>
</tr>
</tbody>
</table>

We divided the images into three groups. Since the number of liver patches are limited, one would want to make sure each group contains at least one image with a large number of liver patches. There are four images with large number of liver patches ({3, 6, 8, 9}). We first randomly chose three out of these four images and put them in separate groups. The remaining 6 images were randomly distributed among the groups (2 for each group). As can be seen from the below calculations, 360 different groupings were possible.

Choosing 3 out of 4 images with
large number of liver patches
Choosing 2 out of the remaining
images for each group respectively

\[
N = C(4,3)C(6,2)C(4,2)C(2,2) = 360,
\]

where \(N\) is the total number of possible groupings. Finally, we selected the grouping that minimized the following cost:

\[
Cost = \sum_{i=1}^{3} \sum_{j=1}^{3} |n_{kid}^i - n_{kid}^j| + |n_{liv}^i - n_{liv}^j|
\]
where \( n_{class}^i \) denotes the number of patches in image \( i \) belonging to the \( 'class' \). This grouping divided the data set into these groups: \( \{ \{3,4,8\}, \{5,6,7\}, \{1,2,9\} \} \); where each number in the sets denotes the image number. The cross validation was done by selecting one group as test and the other two as training sets at each time (repeating 3 times by putting a different group aside as test) and using only the selected 20 features. We repeated the random forests training 20 times for each test group and measured the mean of accuracy, positive predictive value, negative predictive value, sensitivity, and specificity [21].

The averages of these values among the three different test groups are shown in Table 2. Sensitivity and specificity show the percentage of correctly classified kidney and liver patches, respectively. The lower accuracy of the liver classification is probably due to the unbalanced data set. Adjusting the cut-off parameter as explained previously can improve the specificity at the price of lower sensitivity.

\[
\begin{array}{cccccc}
\text{Accuracy} & \text{Positive predictive value} & \text{Negative predictive value} & \text{Sensitivity} & \text{Specificity} \\
0.94 & 0.97 & 0.71 & 0.96 & 0.79
\end{array}
\]

### Table 2: Cross validation classification results; kidney is considered positive and liver negative in definitions.

**CONCLUSIONS AND FUTURE WORK**

US image interpretation, including tissue characterization, is challenging due to variability in quality of US images and in acquisition protocols. Therefore, an objective and quantitative US tissue characterization tool can help in the clinical analysis and interpretation of radiologic data by the clinicians. In this paper, we proposed a tissue characterization technique for in-vivo 3D US images based on the distribution of US backscatter data. We performed feature selection using random forests and showed 94% accuracy in the overall classification of liver and kidney tissue based on selected features, including a combination of homodyned-K parameters and speckle correlation coefficients. We plan to use the proposed feature selection and classification technique to achieve the automatic segmentation of liver and kidney in 3D US data. Future work will also focus on the detection and characterization of hydronephrosis from US images to reduce the use of imaging protocol with ionizing radiation on pediatric patients. The proposed technique may also be applicable to the characterization of other types of soft tissue in US image data.

**ACKNOWLEDGEMENTS**

The authors would like to thank David Hruska for providing the RSK algorithm codes. This project was supported by a philanthropic gift from the Government of Abu Dhabi to Children’s National Medical Center.

**REFERENCES**