Tissue Identification in Ultrasound Images using Rayleigh Local Parameter Estimation

Santiago Aja-Fernández, Marcos Martin-Fernández and Carlos Alberola-López*
LPI, ETSI Telecomunicación, Universidad de Valladolid, SPAIN
sanaja@tel.uva.es, marcma@tel.uva.es, caralb@tel.uva.es

Abstract

A method to identify different tissues in ultrasound images is presented. Assuming a Rayleigh model for speckle, the different tissues present in the image will be related to the Rayleigh σ parameter, with different values for each different tissue. The parameter is locally estimated using well known estimation methods for the Rayleigh distribution, such as the Maximum Likelihood estimator. A significant increase in the separability of the different tissues present in the image is then achieved. A subsequent clustering process based on the position of the maxima of the estimator distribution allows the classification of each pixel in a tissue class.

1. Introduction

Speckle is a kind of granular noise that can be found in many types of coherent imaging systems, such as synthetic aperture radar (SAR), laser illuminated or ultrasound images. Different statistical models have been proposed to deal with such a noise, based mainly on different assumptions. In this paper we will assume a large number of scatters, the phases of the signal are random and independent of the amplitudes, and there is no periodicity in the disposition of the scatters. This way, we will be able to model the envelope as a Rayleigh distribution [3]. This distribution is parametrized by σ, which is related to the characteristics of the original image (without noise), and therefore to the characteristics of the underlying biological tissue that produces the image. This property can be employed for the automatic analysis of ultrasound images for computer-aided diagnosis.

In this work we will assume that the different tissues in ultrasound images will have a uniform value of σ, different for each tissue. Assuming also a Rayleigh distribution of the speckle, we will be able to estimate those values using well-known estimators for σ for data following a Rayleigh distribution. This estimators will be calculated locally. The distributions of the proposed estimators have the feature of being centered in the same place of the original Rayleigh data, but with a smaller variance. This local estimation allows a better identification of regions characterized by different σ parameter and, consequently, belonging to different type of tissue.

2. Statistical Models for Speckle

Several models have been proposed to deal with the output of non-coherent detectors of ultrasonic signals. Considering the received signal, if the phases are random and independent of the amplitudes and the number of scatters is very large and there is no periodicity in the disposition of the scatters, the envelope X may be modelled as a Rayleigh distribution [3]:

\[ p(X) = \frac{X}{\sigma^2} e^{-\frac{X^2}{2\sigma^2}} u(X) \]

where \( u(\cdot) \) is the Heaviside step function and \( \sigma \) is the reflectivity component which comes from the standard deviation of the in-phase and quadrature components of the received signal which are known to be Gaussian distributed. If a specular component is added to the received signal given, the distribution of the envelope becomes a Rician distribution [10]. When in the Rayleigh model a random fluctuation on the number of scatters is allowed we come up with the K distribution for the envelope [4]. Finally a general model is met by merging the Rice and the K distribution, which is referred to as homodyned K distribution [5]. Other models have also been proposed.

It is an usual task to reduce the dynamic range of a ultrasound image to make it match in the smaller dynamic range of a display device. This is done by means of a logarithmic compression, which change the PDF of the different models...
previously reported. In this paper we will assume that there is no logarithmic compression in our data, or more generally, that this logarithmic compression has been inverted.

3. Local Statistics of the Rayleigh distribution

According to Section 2, the Rayleigh distribution is one of the ways to statistically model speckle in an image. This distribution usually arises when taking the absolute value of normal independent random variables with same variance \( \sigma \) [9]. The main parameters of this distribution are as follows:

- **Mean**: \( E\{x\} = \sigma \sqrt{\frac{2}{\pi}} \)
- **Median**: \( \text{median}\{x\} = \sigma \sqrt{\log(4)} \)
- **Mode**: \( \text{mode}\{x\} = \sigma \)
- **Variance**: \( \text{Var}\{x\} = 2 \sigma^2 \)
- **Second order moment**: \( E\{x^2\} = 2 \sigma^2 \)

Assuming that \( R_i(\sigma^2), i = \{1, \cdots, N\} \) is a set of random variables with Rayleigh distribution, two estimators for \( \sigma \) may be defined [8]:

- **Maximum Likelihood (ML) estimator of \( \sigma^2 \)**:
  \[
  \hat{\sigma}_{ML}^2 = \frac{1}{2N} \sum_{i=1}^{N} R_i^2 = \frac{1}{2} \left( \frac{R_i}{\sigma} \right)^2
  \]
  (2)

- **An unbiased estimator of \( \sigma \), according to [8]**:
  \[
  \hat{\sigma}_u = \sqrt{\frac{2}{\pi N}} \sum_{i=1}^{N} R_i = \sqrt{\frac{2}{\pi}} \left( \frac{R_i}{\sigma} \right)
  \]
  (3)

whose variance is larger than the Cramer Rao Lower Bound. Following we will study the distribution of some of this two estimators.

The sample local second order moment of a Rayleigh distribution follows a Gamma distribution \( \gamma(\alpha, \beta) \) [7] with parameters \( \alpha = N \) and \( \beta = \frac{\pi}{2} \sigma^2 \). The mean of such distribution is \( E\{S\} = \alpha \beta = 2 \sigma^2 \) and the mode \( \text{mode}\{S\} = (\alpha - 1) \beta = N \sigma^2 / 2 \). So, if \( N >> \), we can say that \( E\{S\} \approx \text{mode}\{S\} \), i.e., the mean and the mode are placed in the same value. The variance of such a distribution is \( \text{Var}\{S\} = \alpha \beta^2 = 4 \sigma^4 / N \), which decreases when the number of samples grows.

Let us now focus on the sample mean of the image. Let \( R_i(\sigma^2), i = \{1, \cdots, N\} \) be a set of random variables with Rayleigh distribution and

\[
S = \frac{1}{N} \sum_{i=1}^{N} R_i(\sigma^2)
\]

The sum of Rayleigh distributions is a classical hard-to-find problem in communications. Some approximations are usually employed [2]. The mean is given by \( E\{S\} = \sigma \sqrt{\frac{2}{\pi}} \). Depending on the simplifications made [1], the mode can be approximate when \( N >> \) by \( \text{mode}\{S\} \approx \sigma_n \sqrt{\frac{2}{\pi}} \). Again, if \( N >> \) the mean and the mode are approximately equal. Assuming that the variables are independent and uncorrelated, the variance of the sample mean distribution will be \( \text{Var}\{S\} = \frac{1}{N} \text{Var}\{R(a)\} = \frac{1}{N} \sigma^2 2^{\frac{k}{2}} \), which decreases when the value of \( N \).

4. Tissue identification in ultrasound images

As we have already seen, a Rayleigh distribution is parametrized by \( \sigma \), which in ultrasound images is related to the characteristics of the biological tissues. As a starting point we will assume that each tissue present in the image will have very similar values of this parameter, and different tissues will have different \( \sigma \) values. As exposed in Section 3, some parameters in the Rayleigh distribution, like the mean, the median and the mode, are proportional to the value of \( \sigma \), while the variance and the second order moment are proportional to \( \sigma^2 \). Thus, the local estimation of such statistics allows the identification of regions characterized by different \( \sigma \) parameters and, consequently, belonging to different types of tissues. This property can be employed for the automatic analysis of ultrasound images for computer-aided diagnosis.

Assume a set of Rayleigh data \( R_i(\sigma_i) \), with different values of the parameter \( \sigma_i, i = \{1, \cdots, M\} \). As an example, in Fig. 1-(a) a synthetic image with three values of \( \sigma \) is shown. The speckle image generated using a Rayleigh model is in Fig. 1-(b). We can estimate the PDF of such images through their histograms. These histograms are shown in Fig. 2. The histogram of the original image is simply composed by three impulses in the values of the uniform areas, i.e. the

![Figure 1. Synthetic image example. Three regions with uniform \( \sigma \) are supposed.](image)
original values of $\sigma_i$. The histogram of the speckle image is the sum of Rayleigh histograms. Note that the value of $\sigma_i$ may be easily find (first impulse on the left), but not the other two values, due to some overlapping.

Assuming that inside a small region $\eta_{ij}$ around the pixel $ij$ the data may be modelled as Rayleigh with only one $\sigma_i$ value, we can estimate this value using one of the estimators in eq. (2) or in eq. (3). This assumption will fail in the borders. If multiple samples of the same image are available, the statistics may be calculated along the different samples, without using any neighbourhood.

Thus, an estimation of $\sigma$ is then made for each of the pixels in the image, obtaining a new image $I_{\hat{\sigma}_{ij}}$. If eq. (3) is used, and being $I$ the speckle image, then

$$I_{\hat{\sigma}_{ij}} = \sqrt{\frac{2}{\pi} \frac{1}{|\eta_{ij}|}} \sum_{p \in \eta_{ij}} I_p$$

The ML estimator of eq. (2) may have been used instead. As the median of a Rayleigh distribution is also related to the $\sigma$ parameter, we could think in proposing another local estimator $I_{\hat{\sigma}}$ as

$$I_{\hat{\sigma}_{ij}} = \sqrt{\frac{1}{\log(4)}} \text{median}\{I_p\}$$

This new estimator will be tested later.

In Fig. 1-(c) image $I_{\hat{\sigma}}$ — when using eq. (3) and $7 \times 7$ neighbourhood — is shown. Notice that each of the pixel of such image has a follows a distribution proportional to eq. (2)\footnote{Proportional and not equal due to the $\frac{1}{\log(4)}$ factor in eq. (4).}. In Fig. 2 its histogram is depicted. Notice that this histogram is multimodal, and the maxima are placed in the same value than the original $\sigma_i$ values.

Although the mean of the Rayleigh distribution and the distribution of the sample mean of Rayleigh variables is the same, the variance of the latter is smaller. Assuming that the variables are independent and uncorrelated, as we have seen in Section 3 the variance of the sample mean distribution will be $\text{Var}\{S\} = \frac{1}{|\text{\eta}_{ij}|} \text{Var}\{R(\sigma)\}$, being $\text{Var}\{R(\sigma)\}$ the variance of the Rayleigh distribution. Accordingly, an important increase in the separability of the different modes present in the image is then achieved: the different distributions in the histogram in Fig. 2 may be better differentiated for $I_{\hat{\sigma}}$ than for $I$.

The values of the different $\sigma_i$ may be estimated using the maxima of the histogram

$$\hat{\sigma} = \arg \max_{i \in \mathbb{N}} \{p(I_i)\}$$

being $\max_{j=1}^{\infty}$ the value of the $i^{th}$ mode, and $p(I_i)$ the distribution of $I_i$.

If using the ML estimator:

$$\hat{\sigma} = \frac{1}{2} \arg \max_{i \in \mathbb{N}} \{p(I^2)\}$$

The value of the mean and the mode in both distributions is equivalent. So, the centroids may be identified using any of both parameters.

Once the maxima are estimated, the value of each pixel in the image is to be assigned to one of these $\hat{\sigma}$ values. One easy way to do that is using some clustering method, as the well-known K-means [6] (though more complex methods may be used). K-means is an algorithm to cluster objects, based on attributes, into $k$ partitions, attempting to find the centroids of natural clusters in the data. To do that, the algorithm minimize the intra-cluster variance through a squared error function.

Using K-means with 3 clusters directly over image in Fig. 1-(c) will give as result the image in Fig 1-(d). Note that the original tissues are perfectly identified. However, there is some distortion in the edges, due to an effect of the windows used to estimate the local parameters. This problem may be overcome by using multiple samples of each image, or by using anisotropic estimators.

5. Experiments

Some experiments are carried out to illustrate the behaviour of the proposed algorithm.

A synthetic image with 4 different tissue values has been generated, as shown in Fig. 3-(a). A simulated speckle image has been created using a Rayleigh model, as shown in Fig. 3-(b). The histogram of both images are depicted in Fig. 5. The original image has only 4 possible values, corresponding to the four different synthetic tissues. The speckle image histogram (solid line) is the sum of the different Rayleigh distributions. As a result, the original values are now hidden. If we attempt to to run a clustering method straight over this speckle image, it will not be able to detect the different modes. Specifically we run a K-means algorithm with 4 centroids, and results are in Fig. 3-(c). Due to the noise the algorithm has not been able to differentiate between tissues.
The method have been used. The local statistics have been calculated along the samples. Although borders are better defined than using square neigbourhoods, there are still some noisy pattern present inside some tissues. Adding some spatial coherence to the algorithm or using some morphological operator would solve this problem.

The tissue identification method will now be used over real ultrasound images. The proposed method is to be apply in an application to achieve the automatic diagnosis of the developmental dysplasia of the hip (DDH). Starting from the 3D ultrasound volume —from Voluson 730, General Electrics (kretz V730)— of the hip joint the estimation of $I_0$ is done with four different methods: Using the estimator based on the local mean of eq. (3) over the whole volume ($3 \times 3 \times 3$ neigbourhood); estimator based on the local mean over each 2D slice ($7 \times 7$); estimator based on the local median of eq. (5) over each 2D slice, ($7 \times 7$); using the ML estimator in eq. (2) over each 2D slice ($7 \times 7$). The different regions are extracted from the $I_0$ image simply by applying a K-means clustering algorithm with 6 centroids. In Fig. 6 the results of the different methods are shown. In all the cases there is a clear differentiation of the tissues. When using the information of the whole volume, the border effects are smaller, due to the size of the window. Although it has been visually validated by experts, a future numerical validation is to be done.

Finally, the segmented volume has been rendered using thresholding and the marching cubes algorithm, as shown in Fig. 7. The surface in Fig. 7-(a) comes from the original noisy data, and it has been manually thresholded in order to achieve the better results. The surface in Fig. 7-(b) comes from the data after clustering, and the thresholding has been automatically made, as only 6 different values were present.
We make the maxima of the distribution of the localestimator 

time improves significantly the results of the original data 

Although Fig. 7-(b) lacks of connectivity between regions, 

Figure 6. Experiments with a 3D ultrasound volume form the hip joint. One slice of data is shown.

Figure 7. Rendering of the 3D ultrasound data of the hip joint.

6. Discussion and conclusions

A method to increase the separability of the different tissues present in the ultrasound image is proposed. It is based on the local estimation of the \( \sigma \) parameter in a Rayleigh distribution. Thus, the different tissues in an image may be easily recognized and segmented. As we have shown, the use of local estimators, as the MLE, the unbiased estimator based on the local mean, or the proposed estimator based on the median, goes with a significant increase in the separability of the different regions inside the image, and therefore make any subsequent segmentation easier to achieve.

In this paper the identification of the tissues has been done using a simple clustering algorithm, K-means, though more complex algorithms may be used, taking into account that the maxima of the distributions of the local estimators are related with the values of the parameter \( \sigma \). The results here presented may be used as a starting point for more complex segmentation processes in images with speckle. We have shown that the separability of the different tissues is larger when applying some local estimator of \( \sigma \). The resulting image—after or before the clustering process—may be used as starting information for snakes or level set segmentation algorithms.

This work is just a starting point in tissue identification using statistical models in speckle. Some future work still to done as adding spatial coherence to the clustering method, taking into account the artifacts introduced by the patient movement and extending the results to models based on Rician, \( K \) and Homodyned \( K \) distributions. It will be also a future task to deal with the logarithmic compression of the data. The proposed algorithm is to be used in next future in more complex processes to achieve the automatic diagnosis of the developmental dysplasia of the hip (DDH).

References


