

Statistical Analysis of the Envelope of Three-dimensional High-frequency Ultrasonic Backscatter from Dissected Human Lymph Nodes

Thanh M. Bui^{1,2}, Alain Coron^{1,2†}, Jonathan Mamou³, Emi Saegusa-Becroft⁴, Tadashi Yamaguchi⁵, Eugene Yanagihara⁴, Junji Machi⁴, S. Lori Bridal^{1,2}, and Ernest J. Feleppa³

(¹UPMC Univ Paris 06, UMR 7623, LIP, Paris, F-75005 France, ²CNRS, UMR 7623 Laboratoire d'Imagerie Paramétrique, Paris, F-75006 France, ³F. L. Lizzi Center for Biomedical Engineering, Riverside Research, New York, NY, ⁴University of Hawaii and Kuakini Medical Center, Honolulu, HI, ⁵CFME, Chiba University, Chiba, Japan)

1. Introduction

Three-dimensional (3D) high-frequency quantitative ultrasound (QUS) approaches, which are used to characterize tissue microstructure, have previously been demonstrated to reliably determine the presence or absence of metastases in freshly-excised lymph nodes (LNs) of patients with histologically proven cancers [1,2]. The key QUS tasks in evaluating LNs are 3D segmentation and QUS-parameter estimation [2]. However, our current, semi-automatic, 3D segmentation method requires visual inspection and manual correction. The aim of this study is to identify the probability density function (PDF) that best models the envelope of high-frequency ultrasonic data acquired in lymph node tissue. Information regarding this PDF can then be used to develop a more-robust segmentation method based on differences in regional statistical properties [3].

2. Methods

LN dissection, histologic preparation, and ultrasound data-acquisition protocols have been described for this study in [1]. Ultrasonic data acquisition was performed using a focused, single element transducer (PI30-2-R0.50IN, Olympus NDT, Waltham, MA) with an aperture of 6.1 mm, a focal length of 12.2 mm, a center frequency of 25.6 MHz, and a minus-6-dB bandwidth that extends from 16.4 to 33.6 MHz. The theoretically predicted axial and lateral resolutions of the imaging system were 86 and 116 μm , respectively. Radiofrequency (RF) echo signals were digitized at 400 MS/s using an 8-bit Acqiris DB-105 A/D board (Acqiris, Monroe, NY). This study was carried out using a database of 99 LNs dissected from colorectal-cancer patients. Eighteen nodes were almost entirely cancerous and 71 nodes were entirely devoid of cancer.

Following RF digitization, a Hilbert transform was employed to obtain envelope data. For statistical evaluation, non-overlapping

randomly-located cylindrical analysis regions (ARs) with a 0.7-mm length and diameter were extracted from the envelope data. The number of independent resolution cells for each 3D AR was $[\pi*(700/2)^2*700]/[\pi*(116/2)^2*86] \sim 296$. The centers of those ARs were located at the focal depth (12.2 mm). To mitigate a bias related to LN size, a maximum of ten ARs were selected for each LN. Therefore, the total number of analyzed ARs was 962 because some very small LNs provided fewer than 10 ARs.

To determine which PDF best modeled the envelope data, a family of eight exponential PDFs was considered: Generalized Gamma (GG), Generalized Extreme Value (GE), Gamma (GA), Weibull (WE), Loglogistic (LL), Lognormal (LN), Nakagami (NA) and Rayleigh (RA). A maximum-likelihood algorithm was used to estimate fit parameters for each PDF. As an example, the GG PDF is presented in Eqn. (1), where \mathbf{a} and \mathbf{c} are the two shape parameters, and \mathbf{b} is a scale parameter. An interesting property of the GG PDF is its ability to model amplitude and intensity fluctuations. This PDF also includes the following PDFs as special cases: RA ($\mathbf{c}=2$, $\mathbf{a}=1$), Exponential ($\mathbf{c}=1$, $\mathbf{a}=1$), NA ($\mathbf{c}=2$), WE ($\mathbf{a}=1$), GA ($\mathbf{c}=1$) and LN ($\mathbf{a} \rightarrow \infty$).

$$f(x; \mathbf{a}, \mathbf{b}, \mathbf{c}) = \frac{c x^{a c - 1}}{b^{a c} \Gamma(a)} e^{-\left(\frac{x}{b}\right)^c}, \quad x \geq 0 \quad (1)$$

To quantitatively evaluate the goodness of fit of each candidate PDF to the experimental envelope PDF, the Kolmogorov Smirnov (KS) metric [4] was used. Smaller values of the metric indicate a better fit to the experimental envelope PDF.

3. Results

Table 1 indicates that the GG and GE PDFs best model the statistics for 65.28% and 22.15% of the ARs, respectively. The GA PDF best models the statistics for 10.08% of the ARs, but it outperforms all remaining two-parameter PDFs (LL, WE, LN,

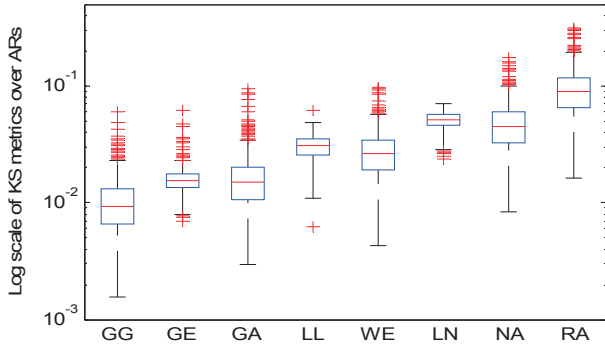


Fig. 1. KS metrics for all ARs in the 71 non-cancerous LNs

Table I. KS metrics for ARs of 99 LNs (#Par.: number of parameters of the PDF; P.Best: percentage of best fit)

PDFs	GG	GE	GA	LL	WE	LN	NA	RA
#Par.	3	3	2	2	2	2	2	1
P. Best	65.28	22.15	10.08	2.08	0.1	0.31	0	0
Mean	0.011	0.016	0.019	0.03	0.03	0.05	0.052	0.104

NA) by a large factor. The bottom row of Table I represents the mean KS metrics of the 962 ARs; it indicates that the GG PDF has the lowest mean value as well as providing the best fit for most ARs. **Fig. 1** shows a box plot of the KS metrics of the ARs of the 71 non-cancerous LNs for the eight PDFs. The plot illustrates how the GG PDF best fits the experimental envelope PDF, followed by the GE and then GA PDFs. Among the 18 cancerous LNs, the GG PDF still outperforms the other PDFs, as shown in **Fig. 2**. The GE PDF is reliable with the second-best fits cancerous as well as non-cancerous LNs. The distribution of the KS metrics for the GA PDF is wider for cancerous LNs, but it still models the experimental envelope PDF better than all other two-parameter PDFs.

For each LN, the percentage of ARs that are best represented by the GG versus GE PDF is illustrated in **Fig. 3** to compare the two PDFs. The 18 cancerous LNs are outlined with a dashed rectangular border. As shown in Fig. 3, there are 85 LNs for which the GG PDF better models the statistics of more than 50% of the ARs of each LN. Two cancerous LNs (and none of non-cancerous LNs) get a 0% value, meaning that all their ARs are better modeled by the GE PDF. **Fig. 4** shows the experimental PDF of an AR of a cancerous LN as well as the best four estimated PDFs and the associated KS metrics.

4. Conclusions

This study indicates that the GG distribution best models the envelope data of cancerous and non-cancerous LNs near the focal zone, and that the GA distribution outperforms other two-parameter PDFs in fitting the envelope data. We plan to examine the statistics of the LN envelope before and after the focal zone in regions where the resolution cell is larger and signal-to-noise ratio is

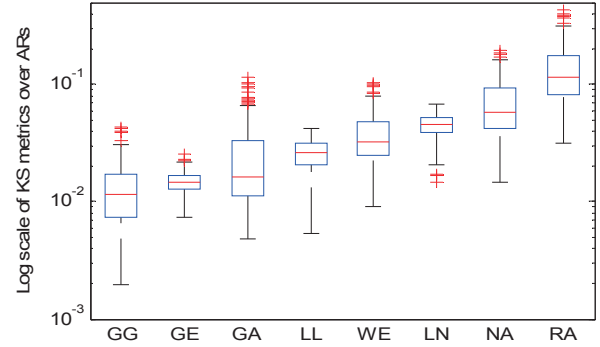


Fig. 2. KS metrics for all ARs in the 18 cancerous LNs

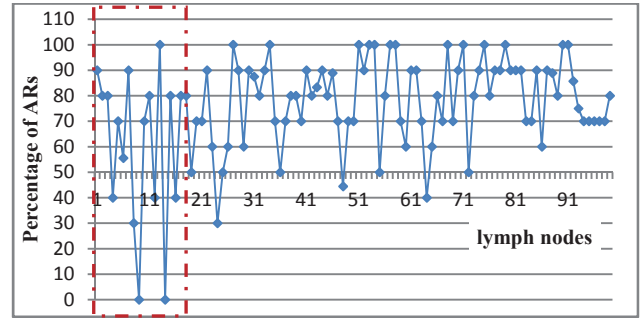


Fig. 3. The percentage of ARs in each LN best represented by GG versus GE distributions

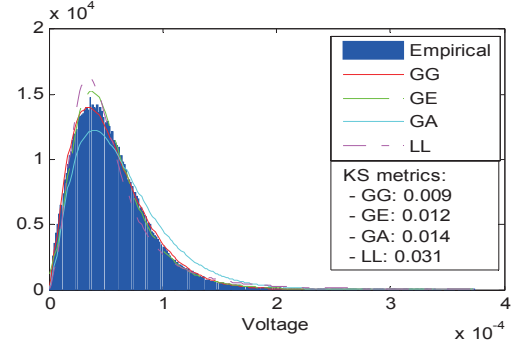


Fig. 4 Estimated PDFs and overlaid fits for the four best-fitting distributions.

lower. By incorporating the statistical properties of the data determined by this study, a more-robust segmentation method can be developed.

Acknowledgments

This research was supported by NIH grant 5R01CA100183 awarded to Riverside Research, Ernest Feleppa, PI.

References

1. J. Mamou, A. Coron, M. L. Oelze, E. Saegusa-Becroft *et al.*, *Ultras. Med. Biol.* **37** (2011) 345.
2. A. Coron, J. Mamou, E. Saegusa-Becroft, M. L. Oelze, *et al.*, *Proc. I.S.B.I. Spain* (2012) 1064
3. J. Anquez, E. D. Angelini, G. Grange and I. Bloch, *IEEE Trans. on Biomed. Eng.*, **60** (2013) 1388.
4. J. D. Gibbon, S. Chakraborti: *Nonparametric Statistical Inference* (Marcel Dekker, Inc 2003) p.111.