

ULTRASONOGRAPHIC PLAQUE CHARACTERIZATION USING A RAYLEIGH MIXTURE MODEL

José Seabra*, João Sanches

Institute for Systems and Robotics
Instituto Superior Técnico
1049-001 Lisboa, Portugal

Francesco Ciompi, Petia Radeva

Computer Vision Center
08193 Bellaterra (Barcelona), Spain

ABSTRACT

A correct modelling of tissue morphology is determinant for the identification of vulnerable plaques. This paper aims at describing the plaque composition by means of a Rayleigh Mixture Model applied to ultrasonic data. The effectiveness of using a mixture of distributions is established through synthetic and real ultrasonic data samples. Furthermore, the proposed mixture model is used in a plaque classification problem in Intravascular Ultrasound (IVUS) images of coronary plaques. A classifier tested on a set of 67 *in-vitro* plaques, yields an overall accuracy of 86% and sensitivity of 92%, 94% and 82%, for fibrotic, calcified and lipidic tissues, respectively. These results strongly suggest that different plaques types can be distinguished by means of the coefficients and Rayleigh parameters of the mixture distribution.

Index Terms— Plaque Composition, Rayleigh Mixture Model, Intravascular Ultrasound

1. INTRODUCTION

Vulnerable plaques are defined as lesions presenting high risk of rupture, possibly leading to brain stroke or heart attack. It is important to objectively characterize the plaque morphology to identify these kind of lesions. Intravascular ultrasound (IVUS) is an imaging technique which allows to clearly assess the arterial wall internal morphology.

Appearance-based methods [1] were pursued to qualitatively characterize plaque morphology as soft (echolucent), fibrous (intermediate echogenicity), mixed (several acoustical subtypes) and calcified (strongly echogenic). Given the high variability in the appearance of tissues, the IVUS imaging parameters are often changed to improve visualization, turning the application of appearance-based methods into a challenging task. Hence, recent studies on plaque composition perform analysis on the raw Radio Frequency (RF) signals [2]. In this paper we use an objective and reproducible method for image reconstruction and tissue assessment from the RF data, avoiding machine- and operator- dependent settings [3].

*This work was supported by Fundação para a Ciência e a Tecnologia (ISR/IST plurianual funding) through the POS Conhecimento Program which includes FEDER funds.

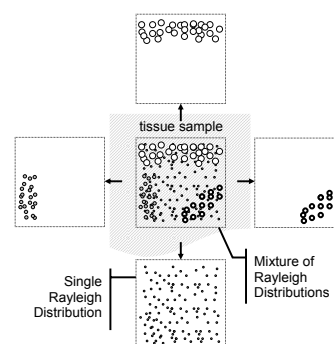


Fig. 1. Tissue acoustic model, where different scattering phenomena may occur.

It is well established that under particular conditions [4], pixel observations in ultrasound images can be modelled by Rayleigh statistics, while deviations to the so-called *fully developed speckle* are better described by other more complex distributions [5]. Since plaque morphology may present different types of components and spatial organization/complexity, the use of such statistical models may be valid but not sufficient. In this paper we propose to use a mixture of Rayleigh distributions, designated as Rayleigh Mixture Model, for a complete description of plaque morphology. This model associates the mathematical easiness of using the Rayleigh model with the robustness of using a mixture of distributions.

2. METHODS

2.1. Tissue Acoustic Model

In medical ultrasound, a transmitted pulse interacts with a certain anatomical region, providing information about inner tissue structures. Here, we hypothesize that the morphology of a scanned tissue sample results from different scattering phenomena, as depicted in Fig. 1. The backscattered signal is thus dependent on the number of scatterers as well as their size. As pointed out in [4] these features can be considered as histological descriptors of tissues.

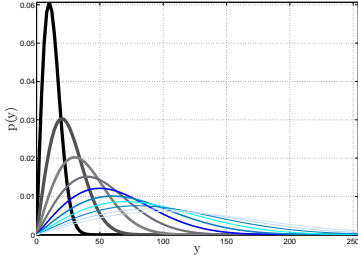


Fig. 2. Rayleigh pdfs, created with $10^2 < \sigma^2 < 10^3$ (from darker to lighter curves.)

2.2. Rayleigh Mixture Model (RMM)

In this section we formulate the RMM and describe the estimation method which provides the mixture coefficients (weights) and the Rayleigh parameters associated with each mixture component (distribution).

We define $\mathbf{Y} = \{y_i\}$ as a set of pixel intensities obtained from a region of interest, specifically a plaque, taken from an ultrasound image. Pixel intensities are considered random variables which are described by the following mixture of L distributions:

$$p(y_i|\Psi) = \sum_{j=1}^L \theta_j p_j(y_i|\sigma_j), \quad (1)$$

where

$$p_j(y_i|\sigma_j) = \frac{y_i}{\sigma_j^2} e^{-\frac{y_i^2}{2\sigma_j^2}} \quad (2)$$

is the Rayleigh probability density function (pdf) parameterized by σ_j . The effect of changing σ in the shape of the distribution and thus in the image intensity is illustrated in Fig. 2. In (1), $\Psi = (\theta_1, \dots, \theta_L, \sigma_1, \dots, \sigma_L)$ are the parameters to be estimated, accounting for the weights θ_j and the Rayleigh parameters σ_j of each mixture component. In this case, the condition $\sum_{j=1}^L \theta_j = 1$ must hold to guarantee that $p(y_i|\Psi)$ is a true distribution function. The Rayleigh parameters σ_j associated with y_i , characterize the acoustic properties of the tissue at the i^{th} location. The joint distribution of the pixels, considered independent and identically distributed (i.i.d), is given by:

$$p(\mathbf{Y}|\Psi) = \prod_i^N p(y_i|\Psi). \quad (3)$$

The goal is to estimate Ψ by maximizing the likelihood function such that:

$$\hat{\Psi}_{ML} = \arg \max_{\Psi} \mathcal{L}(\mathbf{Y}, \Psi), \quad (4)$$

where

$$\mathcal{L}(\mathbf{Y}, \Psi) = \log p(\mathbf{Y}|\Psi) = \sum_{i=1}^N \log \left(\sum_{j=1}^L \theta_j p_j(y_i|\sigma_j) \right). \quad (5)$$

The maximization of (5) is a difficult task because it consists of a logarithmic function of a sum of terms. To overcome this difficulty the *Expectation-Maximization* (EM) [6] method is used where a set of hidden variables are introduced, $\mathbf{K} = \{k_i\}$ where $k_i \in \{1, \dots, L\}$. The value of $k_i = j$ informs us about the mixture component j that generated the i^{th} observation, y_i , with probability $p_{k_i}(y_i|\sigma_{k_i})$ defined in (2). Each n^{th} iteration of the EM method is composed of two steps:

- E step: where the expectation of the new likelihood function, $\mathcal{L}(\mathbf{Y}, \mathbf{K}, \Psi)$, is computed with respect to \mathbf{K} ,

$$Q(\mathbf{Y}, \Psi^n, \Psi) = E_{\mathbf{K}} [\mathcal{L}(\mathbf{Y}, \mathbf{K}(\Psi^n), \Psi)] \quad (6)$$

and

- M step: where a new estimate of Ψ , Ψ^{n+1} , is obtained by maximizing the function Q ,

$$\Psi^{n+1} = \arg \max_{\Psi} Q(\mathbf{Y}, \Psi^n, \Psi). \quad (7)$$

These two steps alternate until convergence is achieved. The new likelihood function is

$$\begin{aligned} \mathcal{L}(\mathbf{Y}, \mathbf{K}, \Psi) &= \log p(\mathbf{Y}, \mathbf{K}|\Psi) = \sum_{i=1}^N \log p(y_i, k_i|\Psi) \\ &= \sum_{i=1}^N \log p_{k_i}(y_i|\sigma_{k_i}) + \log \underbrace{p(k_i|\sigma_{k_i})}_{\theta_{k_i}} \end{aligned} \quad (8)$$

where $p_{k_i}(y_i|\sigma_{k_i})|_{k_i=j}$, defined in (2), is the k_i^{th} component of the RMM and θ_{k_i} is the mixture coefficient associated with the k_i^{th} component. The maximization of (8) is impossible because the hidden variables \mathbf{K} are not known. Therefore, the *expectation* with respect to \mathbf{K} is computed as follows:

$$\begin{aligned} Q(\Psi, \hat{\Psi}) &= E_{\mathbf{K}} [\mathcal{L}(\mathbf{Y}, \mathbf{K}, \Psi)|\mathbf{Y}, \hat{\Psi}] \\ &= \sum_{i=1}^N E_{k_i} [\log p_{k_i}(y_i|\sigma_{k_i}) + \log p(k_i|\sigma_{k_i})] \\ &= \sum_{i=1}^N \sum_{j=1}^L \gamma_{i,j} [\log p_j(y_i|\sigma_j) + \log \theta_j], \end{aligned} \quad (9)$$

where $\hat{\Psi} = (\hat{\theta}_1, \dots, \hat{\theta}_L, \hat{\sigma}_1, \dots, \hat{\sigma}_L)$ is the previous estimation of the parameters and $\gamma_{i,j}$ is the distribution of the unobserved variables which is defined as follows:

$$\gamma_{i,j} = p(k_i = j|y_i, \hat{\Psi}) = \frac{p_j(y_i|\hat{\sigma}_j)p(k_i = j)}{p(y_i|\hat{\Psi})}. \quad (10)$$

In (10), $p_j(y_i|\hat{\sigma}_j)$ is computed by using (2), $p(k_i = j) = \hat{\theta}_j$ and, by definition, $p(y_i|\hat{\Psi}) = \sum_{j=1}^L p_j(y_i|\hat{\sigma}_j)$. The likelihood function (9) can be rewritten by separating the terms

which depend exclusively on θ_j and σ_j , and considering (2) resulting in:

$$\mathcal{Q}(\Psi, \hat{\Psi}) = \sum_{i=1}^N \sum_{j=1}^L \gamma_{i,j} \log(\theta_j) + \sum_{i=1}^N \sum_{j=1}^L \gamma_{i,j} \left[\log\left(\frac{y_i}{\sigma_j}\right) - \frac{y_i^2}{2\sigma_j} \right]. \quad (11)$$

To estimate θ_j , we introduce the *Lagrange multiplier* λ to solve the following equation:

$$\frac{\partial}{\partial \theta_j} \left[\sum_{i=1}^N \sum_{j=1}^L \gamma_{i,j} \log(\theta_j) + \lambda \left(\sum_j \theta_j - 1 \right) \right] = 0, \quad (12)$$

resulting in: $\frac{1}{\theta_j} \sum_{i=1}^N \gamma_{i,j} = -\lambda$. If we sum both sides over j , we get that $\sum_{j=1}^L \lambda_{i,j} = N$ and $\lambda = -N$, finally yielding:

$$\hat{\theta}_j = \frac{1}{N} \sum_{i=1}^N \gamma_{i,j}. \quad (13)$$

The Rayleigh parameter of each mixture component, σ_j , is computed by finding the stationary point of \mathcal{Q} with respect to Σ , $\nabla_{\Sigma} \mathcal{Q} = 0$ such that:

$$\frac{\partial}{\partial \sigma_j} \left[\sum_{i=1}^N \sum_{j=1}^L \gamma_{i,j} \left(\log\left(\frac{y_i}{\sigma_j}\right) - \frac{y_i^2}{2\sigma_j} \right) \right] = 0, \quad (14)$$

which is easily solved for σ_j to obtain:

$$\hat{\sigma}_j = \frac{\sum_{i=1}^N \gamma_{i,j} \frac{y_i^2}{2}}{\sum_{i=1}^N \gamma_{i,j}} = \frac{1}{N} \sum_{i=1}^N \gamma_{i,j} \frac{y_i^2}{2}. \quad (15)$$

The EM algorithm is initialized with uniformly weighted coefficients $\Theta = (\theta_j) = \frac{1}{L}$ while the mixture parameters are assigned with the Maximum Likelihood (ML) estimator, $\hat{\sigma}_{ML} = \sum_{i=1}^N \frac{y_i^2}{2}$, such that $\Sigma = \{\sigma_j\} = \hat{\sigma}_{ML}$. The initial choice of components was set arbitrarily to $L = 10$; however, when $|\sigma_i - \sigma_j| < \epsilon = 1$ with $(i \neq j) = \{1, \dots, L\}$, then $\sigma_i = \frac{\sigma_i + \sigma_j}{2}$ and $\theta_i = \theta_i + \theta_j$, while σ_j and θ_j are padded to zero. This constraint assures stability of the RMM, particularly, for modelling plaque morphology. Our experiments led us to the value $L = 3$ as being a sufficient input value for the RMM algorithm. Thus, after applying the RMM we get a 7-length feature vector which includes 3 mixture coefficients, 3 Rayleigh parameters and the value M which corresponds to the number of non-zero (effective) mixture components.

3. EXPERIMENTAL RESULTS

3.1. Synthetic Ultrasonic Sample

The adequacy of the RMM algorithm to model tissue morphology is first tested using an ultrasonic phantom. The synthetic image consists of one large ellipse with a background intensity (Rayleigh parameter) containing three smaller ellipses featuring two different θ intensities. Details of region

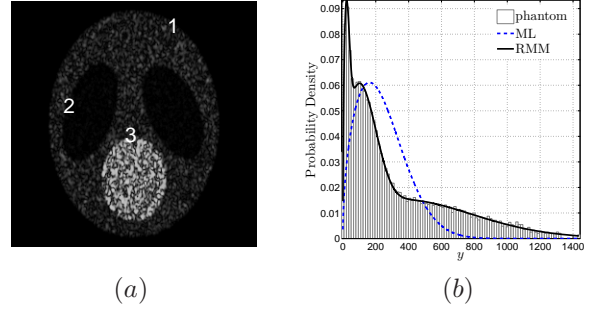


Fig. 3. (a) Phantom image. (b) Pdfs obtained with RMM and ML, overlapped with phantom histogram.

Table 1. RMM estimation results with synthetic image

RMM	Components	Phantom	Estimated
Rayleigh parameters	1	100.0	100.8
	2	20.0	21.3
	3	350.0	357.0
Mixture weights	1	0.59	0.60
	2	0.27	0.26
	3	0.14	0.14

intensities and corresponding proportions (weights) are presented in Table 1. The noisy ultrasonic image (Fig. 3(a)), \mathbf{Y} , is built from the phantom by a convolutional approach, assuming the imaging system has a linear, space-invariant point spread function and a linear transducer according to [7].

This phantom is modelled using the described RMM and the ML estimator of the Rayleigh parameter $\hat{\sigma}$, assuming the traditional single Rayleigh distribution model, according to $\hat{\sigma} = \sqrt{\frac{1}{2nm} \sum_{i,j=1}^{n,m} y_{i,j}^2}$ where the pixels are considered i.i.d. random variables with Rayleigh distribution. In Fig. 3(b) the phantom data histogram is depicted together with the mixture pdf obtained with the RMM and the ML estimated single Rayleigh pdf. It is evident that a single distribution is not able to correctly describe the whole tissue sample. On the other hand, the RMM is able to model the entire tissue sample, providing correct estimates of the coefficients θ_k and σ_k , as shown in Table 1.

3.2. Plaque Composition

The characterization of plaque composition is based on an *in-vitro* IVUS study of plaques obtained from post-mortem human coronary arteries by a Galaxy II IVUS Imaging System (Boston Scientific) with a catheter Atlantis SR Pro 40 MHz (Boston Scientific). For each artery, RF data is acquired at different cross-sections, and then processed according to a rigorous reconstruction protocol [3] to obtain images corresponding to the envelope of the RF data in cartesian coordinates (Fig. 4(a)). It is worth to note that, for what concerns the image properties, these processing operations do not change the statistical properties of the data, specifically, the ability to be modelled with Rayleigh pdfs. *In-vitro* data was sampled from 8 coronary specimens, resulting in 45 IVUS images. Plaques were labelled by an expert interventionist and a

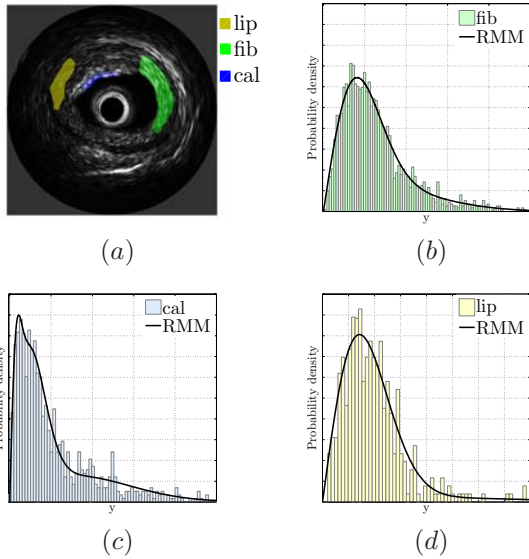


Fig. 4. (a) IVUS image containing 3 distinct labelled plaques. (b-d) Mixture pdfs estimated with RMM.

Table 2. Mean values of Rayleigh parameters and Mixture coefficients estimated with RMM applied for the dataset of 67 plaques

RMM	Components	lipidic	fibrotic	calcified
Rayleigh parameters	1	188	140	318
	2	410	275	1171
	3	-	555	3390
Mixture coefficients	1	0.82	0.51	0.33
	2	0.18	0.39	0.46
	3	-	0.10	0.21

pathologist, according to histology, resulting in 24 fibrotic, 12 lipidic and 31 calcified plaques (example in Fig.4a). We first investigate how the RMM behaves for different tissue types. Hence, we consider an IVUS example containing three different tissue types. It is observed in Fig. 4(b)-Fig. 4(d) that the RMM is a robust ultrasonic modelling method since it describes correctly each tissue type. Additionally, the application of the RMM to our dataset determines its usefulness for plaque characterization because it provides specific outcomes according to each tissue type (Table 2).

The estimated RMM parameters are used as discriminative features in a multi-class characterization framework, where the Error Correcting Output Code (ECOC) technique [8] is employed together with the AdaBoost classifier [9]. The classifier performance is assessed by means of the *Leave-One-Patient-Out* (LOPO) [3] cross-validation technique, where the training set, containing 7 RMM features per plaque (3 mixture coefficients, 3 Rayleigh parameters and the number of effective mixture components) is built taking at each time all *in-vitro* cases, except one, used as test. The RMM estimation for each plaque is performed by considering all the pixels enclosed in it. Classification results are

Table 3. Performance results. Sensitivity: $S = \frac{TP}{TP+FN}$, Specificity $K = \frac{TN}{TN+FP}$, global Accuracy $A = \frac{TP+TN}{TP+TN+FP+FN}$, where TP = true positive, TN = true negative, FP = false positive and FN = false negative

LOPO	S_{fib}	S_{cal}	S_{lip}	A
median	65.00 (39.09)	81.53 (20.34)	44.00 (37.82)	66.30 (15.92)
ML	41.67 (46.85)	0.00 (0.00)	90.42 (15.84)	44.17 (36.28)
RMM	91.67 (13.94)	93.75 (15.30)	82.00 (24.90)	85.56 (18.85)

presented in Table 3. It becomes clear that the application of RMM outperforms the classification results obtained using solely the ML Rayleigh parameter or the median intensity. Regarding the classification obtained with RMM, it is observed that lipidic plaques are well described by 2 mixture components, while calcified and fibrotic plaques are modelled by 3 components, with difference in the range of estimated Rayleigh parameters (Table 2).

4. CONCLUSIONS

This paper presented an algorithm based on mixture of Rayleigh distributions to model and characterize the tissue morphology in ultrasonic data. The application of the RMM in both synthetic and real data demonstrates the adequacy of a mixture of distributions to model ultrasonic data. The usefulness of RMM features for tissue characterization was established through a practical plaque characterization problem in IVUS, with high classification scores being achieved.

5. REFERENCES

- [1] E. Gussenhoven, C. Essed, and P. Frietman, "Intravascular ultrasonic imaging: histologic and echographic correlation," *European Journal of Vascular Surgery*, vol. 3, pp. 571–576, January 1989.
- [2] A. Nair, B. Kuban, E. M. Tuzcu, P. Schoenhagen, S. Nissen, and D. Vince, "Coronary plaque classification with intravascular ultrasound radiofrequency data analysis," *Circulation*, vol. 106, pp. 2200–2206, 2002.
- [3] K. L. Caballero, J. Barajas, O. Pujol, O. Rodriguez, and P. Radeva, "Using reconstructed ivus images for coronary plaque classification," in *EMBS, Cité Internationale, Lyon, France, 2007*, pp. 2167–2170.
- [4] J. Thijssen, "Ultrasonic speckle formation, analysis and processing applied to tissue characterization," *Pattern Recognition Letters*, vol. 24, no. 4-5, pp. 659–675, 2003.
- [5] T. Eltoft, "Modeling the amplitude statistics of ultrasonic images," *IEEE Transactions on Medical Imaging*, vol. 25, no. 2, pp. 229–240, Feb 2006, Comparative Study.
- [6] Arthur P. Dempster, Nan M. Laird, and Donald B. Rdin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B*, vol. 39, pp. 1–38, 1977.
- [7] C. Butakoff, S. Balocco, S. Ordas, and A. F. Frangi, "Simulated 3-d ultrasound lv cardiac images for active shape model training," in *Medical Imaging 2007: Image Processing*, 2007.
- [8] T. G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *Journal of Artificial Intelligence Research*, vol. 2, pp. 263–286, 1995.
- [9] Robert E. Schapire, "Using output codes to boost multiclass learning problems," in *Proc. 14th International Conference on Machine Learning*, 1997, pp. 313–321.