Linear classifier and textural analysis of optical scattering images for tumor classification during breast cancer extraction

Alma Eguizabal[a], Ashley M. Laughney[bd], Pilar Beatriz Garcia Allende[c], Venkataramanan Krishnaswamy[d], Wendy A. Wells[e], Keith D. Paulsen[e], Brian W. Pogue[d], Jose M. Lopez-Higuera[a], Olga M. Conde[a]

[a]Univ. de Cantabria (Spain);
[b]Massachusetts General Hospital/Harvard Medical School (United States);
[c]Helmholtz ZentrumMünchen GmbH (Germany)
[d]Thayer School of Engineering at Dartmouth (United States);
[e]Dartmouth Hitchcock Medical Ctr. (United States);

ABSTRACT

Texture analysis of light scattering in tissue is proposed to obtain diagnostic information from breast cancer specimens. Light scattering measurements are minimally invasive, and allow the estimation of tissue morphology to guide the surgeon in resection surgeries. The usability of scatter signatures acquired with a micro-sampling reflectance spectral imaging system was improved utilizing an empirical approximation to the Mie theory to estimate the scattering power on a per-pixel basis. Co-occurrence analysis is then applied to the scattering power images to extract the textural features. A statistical analysis of the features demonstrated the suitability of the autocorrelation for the classification of not-malignant (normal epithelia and stroma, benign epithelia and stroma, inflammation), malignant (DCIS, IDC, ILC) and adipose tissue, since it reveals morphological information of tissue. Non-malignant tissue shows higher autocorrelation values while adipose tissue presents a very low autocorrelation on its scatter texture, being malignant the middle ground. Consequently, a fast linear classifier based on the consideration of just one straightforward feature is enough for providing relevant diagnostic information. A leave-one-out validation of the linear classifier on 29 samples with 48 regions of interest showed classification accuracies of 98.74% on adipose tissue, 82.67% on non-malignant tissue and 72.37% on malignant tissue, in comparison with the biopsy H&E gold standard. This demonstrates that autocorrelation analysis of scatter signatures is a very computationally efficient and automated approach to provide pathological information in real-time to guide surgeon during tissue resection.

Keywords: breast tumor; localized backscattering; scattering power; texture analysis; linear classifier.

1. INTRODUCTION

Breast cancer is the most frequent cancer and the more deaths-related between women worldwide [1], although the survival-incidence rate is quite optimistic due to the good prospective of early stage detection. Breast conserving therapy (BCT) is the standard of care for patients with early and advanced disease (stages I to III) [2], whereas its major limitation is the current inability to asses accurate tumor margins. Margin assessment is routinely conducted post-operatively by standard histology and an intraoperative alternative lacks standardization. BCT has been demonstrated to be as safe as mastectomy when surgical margins are clear of disease [3]. An optimized intraoperative imaging system to discriminate between cancer and normal tissue could substantially reduce the psychological effects of mastectomy or repeated surgery, as well as the risk and costs of reappearance of the disease, which causes a significant detriment in patient prognosis. Optical scattering spectroscopy is an effective intraoperative alternative to routine paraphine histopathology [4]. This paper explores the spatial distribution of tissue scattering and how its natural heterogeneity relates to tissue morphology, which in turn generates different textures in the collected scattering images. A scanning in situ spectroscopy platform is used to sample the reflectance response of breast cancer specimens, to finally explore textural features to design a fast classifier.

Texture analysis has been previously proven to be an effective morphology descriptor in breast tissue characterization [5-9]. Most work focused on the detection of tumor masses in mammograms [5-6], ultrasound captures [7-8], and histopathology images [9] Optical spectroscopy provides higher resolution than mammography and is additionally less

invasive, while histopathology requires extraction and tissue staining. Texture analysis of spectroscopy data has also been reported for the identification of morphological information of tissue [10], including cancer behavior [11].

We propose textural analysis of light scattering images for an easy and accurate detection of malignancy regions. The diagnostic capability of the textural features allows that malignancy regions are readily identified with a linear threshold classifier. The process of analysis is summarized on Figure 1. Scattering images of fresh tissue specimens excised are initially acquired. Texture analysis of these images consisted in the local calculation of the grey level co-occurrence matrix (GLCM) is per pixel within a sliding window. The autocorrelation of these co-occurrence matrices is the indicator to decide the diagnosis of the pixel under analysis, just deciding whether its value is higher or lower than a threshold.
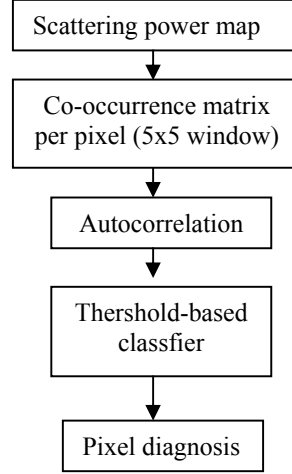

Figure 1. Procedure sequence for tissue diagnosis.

2. MATERIALS AND METHODS

2.1 Localized reflectance spectroscopy

Surgical breast tissue specimens were imaged with a custom-built reflectance system, developed in a previous study [12]. It consists of a confocal spectroscopic set-up and a raster-scanning sample platform built using translation stages. Tissue samples are hydrated with a phosphate buffer solution during the measurement procedure. The system employs a quasi-confocal illumination and detection to constrain the overlapping illumination and detection spot sizes within approximately one scattering distance in tissue (~100 μm in the visible). The optical and electromechanical subsystems are integrated via a custom developed LabVIEW interface. The background response, $R_{bkgrd}(\lambda)$, is subtracted from the measured spectra, $R_{acquired}(\lambda)$, and the data is normalized, $R(\lambda)$, with respect to a diffuse reflectance standard (Spectralon, Labsphere, Inc., North Sutton, New Hampshire), as shown in Eq. 1.

$$R(\lambda) = \frac{R_{acquired}(\lambda) - R_{bkgrd}(\lambda)}{R_{spectralon}(\lambda) - R_{bkgrd}(\lambda)} \tag{1}$$

An empirical approximation to Mie theory is used and a Beer's Law attenuation factor to describe the reflectance, R(λ) [13], as shown on (Eq.2).

$$R(\lambda) = A\lambda^{-b} \exp^{-\Gamma*[HbT]\{StO_2*\varepsilon_{HbO_2}(\lambda)+(1-StO_2)*\varepsilon_{Hb}(\lambda)\}} \tag{2}$$

where Γ refers to the mean optical path-length (dependent on the illumination and detection geometry), $[HbT]$ is the total hemoglobin concentration, StO_2 is the oxygen saturation factor (ratio of oxygenated to total hemoglobin), ε_{HbO_2} and ε_{Hb} refer to the molar extinction coefficients of these two chromophores, respectively (Oregon Medical Laser

Center Database, [14]). *A* and *b* are scattering amplitude and scattering power and both depend on the size and number density of scattering centers in the tissue volume , thereby reflecting variations in breast tissue morphology. The scattering spectrum is relatively featureless compared with absorption spectrum but its spatial distribution is heterogeneous. In a previous publication [15] we verified that the scattering power parameter, *b* , provided better tissue discrimination capability than all other scattering parameters. Consequently, this will be the employed parameter to extract the diagnostic morphology in the present study.

We imaged 29 specimens of breast tissue [13], with 48 different Regions of Interest (ROIs) corresponding to 3 different diagnosis categories: non-malignant, malignant and adipose. All the data is summarized in Table 1.

Table 1. Distribution of the analyzed categories of breast tissue.

Tissue type	No of ROIs
Non-Malignant	25
Malignant	14
Adipose	9
Total ROI	**48**

2.2 Co-occurrence texture analysis

The heterogeneity of the spatial distribution of scattering power allows extracting region-based signatures of diagnostic morphology, such as texture features. In image analysis a texture is defined [16] as the spatial variation in pixel intensities, in this case, the spatial variation of the scattering parameter power values. These features may differ from a tissue types to another, since adipose tissues have large scatterers leading tolower scattering power, whereas fibrous tissues have more cellular and collagen-based structures that lead to higher scattering power [17]. The scattering power can be interpreted as a gray-level variation that differs from one tissue to another, Table 2.

Table 2. Illustration of how textures of different diagnosis look different.

Scattering Power Map	Tissue diagnosis
	Non-Malignant
	Malignant
	Adipose

To be able to define texture properties, the grey-level co-occurrence matrix is employed [16]. This matrix represents the probability of a simultaneous occurrence of a pair of grey pixels in a certain position, combining all possible positions. The matriz *C(k,l;d,φ)* in equation 3 counts for the simultaneous occurrence of pixel k and l with a distance d on a direction give by the angle ϕ. This process is repeated all along the grey values considered, where *g(i,j)* is the grey value of the pair pixel position *(i,j)*. Figure 2 illustrates the meaning of the parameters showed on equation 3.

$$C(k,l,;d,\phi) = \sum\sum \partial(k - g(i,j))\partial(l - g(i + d\cos\phi, j + d\sin\phi)) \tag{3}$$
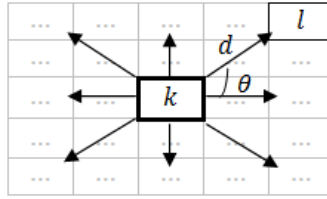
Figure 2. Representation of co-occurrence. The central pixel, k is separated from pixel l with the parameters described on equation 3. The process consists in checking the repetition of this patron along across the whole image, collecting the results in a 2D matrix, as a histogram. This matrix is dubbed grey level co-occurrence matrix (GLCM)

A simplified example of the co-occurrence process is shown on Figure 3. Eight gray level values are considered to map the whole range of scattering power values. Studies performed on the scattering power maps have shown that the computation angle, ϕ, and the distance, d, between pixels are neither determinant nor conclusive in the scattering power textures, so a distance of one pixel and an averaged of different angles have been used. Texture analysis also requires the definition of a spatial domain around the pixel to be able to compute the co-occurrence matrix and their associated metrics. A study about the optimum spatial size [13] lead to a 5x5 pixel window; corresponding to a 500 μm square window. Then, as shown of Figure 4, a sliding window of 5x5 pixels is centered on each pixel to consider the texture of enough tissue to be correctly diagnosed, but not too much so it is safely bounded.
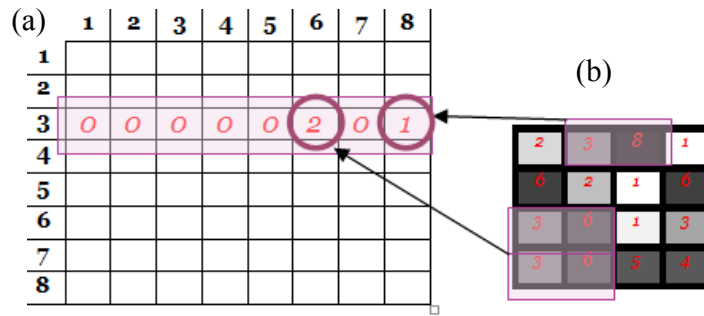

Figure 3. An example of one step in co-occurrence calculation. A 3-bit image (b) with 8 grey levels is represented. The GLCM will be 8x8 to collect all possible combinations. On this particular example the patrons to be searched are pixel value 3 with distance 1 and angle 0, which is third row on the GLCM (a). The result is similar to a 2D histogram.
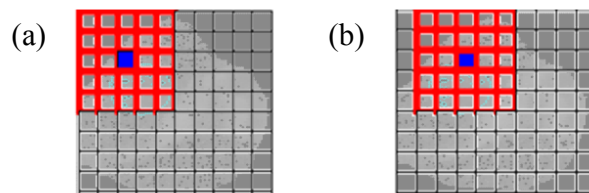

Figure 4. Sliding window to locally calculate texture features. The window of analysis slides from (a) to (b) and in succession, until all image is covered. The resulting texture features of the window correspond to the central pixel.

A GLCM is obtained for each window and it has to be normalized dividing all its elements by the total number of pairs considered. This step is needed to consider the GLCM as a joint probability density, $p(i,j)$. Thereby, several metrics such as the energy, correlation, entropy, homogeneity, autocorrelation, etc. can be obtained from it metrics [16-17]

In this study we compared the diagnostic capability of the textural features mentioned above, which resulted in the autocorrelation exhibiting the most diagnostically power after a statistical analysis of the box plots representation of theses GLMC metrics. This feature provides a measure of gray-tone linear-dependencies in the image [17], and it is defined as:

$$f_1 = \sum_i \sum_j (i\,j)\, p(i,j)$$

(4)

2.3 Linear classifier

Given the diagnostic capability of the autocorrelation, a straightforward classification approach can be based on a linear threshold of this single feature, as illustrated in Figure 5. The malignancy probability is estimated per sample following a leave-one-out method. A leave-one-out method is followed to determine the classification measures of this linear threshold algorithm. It consists in leaving one sample "out", whose malignancy probability is estimated using the other samples as the training set to determine the threshold. This process is repeated over the whole data set, and the classification measures (probabilities of detection and false alarms) are obtained considering the mean of all iterations.
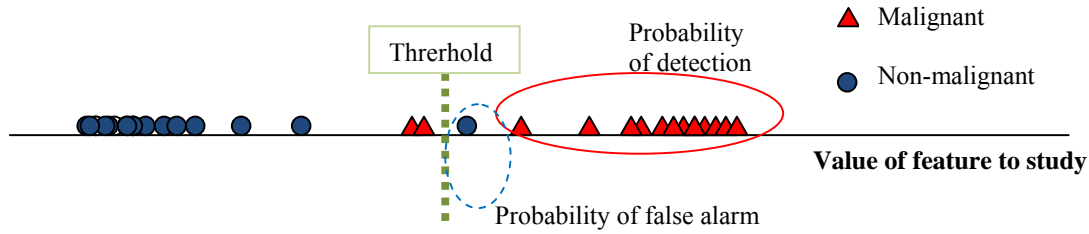

Figure 5. Linear classifier based on the study of just one feature. If a pixel value is higher than a threshold is considered as malignant and if is lower it is not. The choice of this threshold will define the false alarm and the detection rates.

3. RESULTS AND DISCUSSION

Autocorrelation was determined as the GLCM texture feature most suitable for distinguishing normal/benign, malignant and adipose tissue based on the scattering power images. A leave-one-out technique over the 29 different samples allows to define a very constant threshold based on the quartiles of the autocorrelation values per tissue type. Figure 5 shows how this threshold behaves during the leave-one out process and also the box plots of the three different tissue types considered on this analysis.
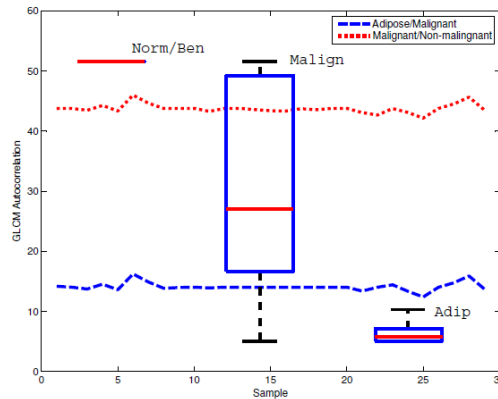

Figure 6. Box plots and threshold variation over the leave-one-out analysis. They are represented all together to see what a linear classifier is based on. Box plots red bars indicate the median per diagnostic class and boxes delineate the interquartile fractions. Dotted red line represents the threshold for malignant/non-malignant discrimination; dashed blue line represents the evolution of the threshold for the adipose/malignant discrimination

The leave-one-out process revealed a mean sensitivity of 98.74% in adipose tissue, 82.67% in non-malignant tissue and 72.37% in malignant tissue, in comparison with the biopsy gold standard. When observing the differences between b, scattering power, itself and autocorrelation of GLCM, the latter shows more contrast to differentiate diagnosis, as shown on Figure 7. Non-malignant tissue shows higher autocorrelation values while adipose tissue presents a very low autocorrelation on its scatter texture, being malignant the middle ground. Consequently, a fast linear classifier based on the consideration of just one straightforward feature is sufficient for providing relevant diagnostic information.
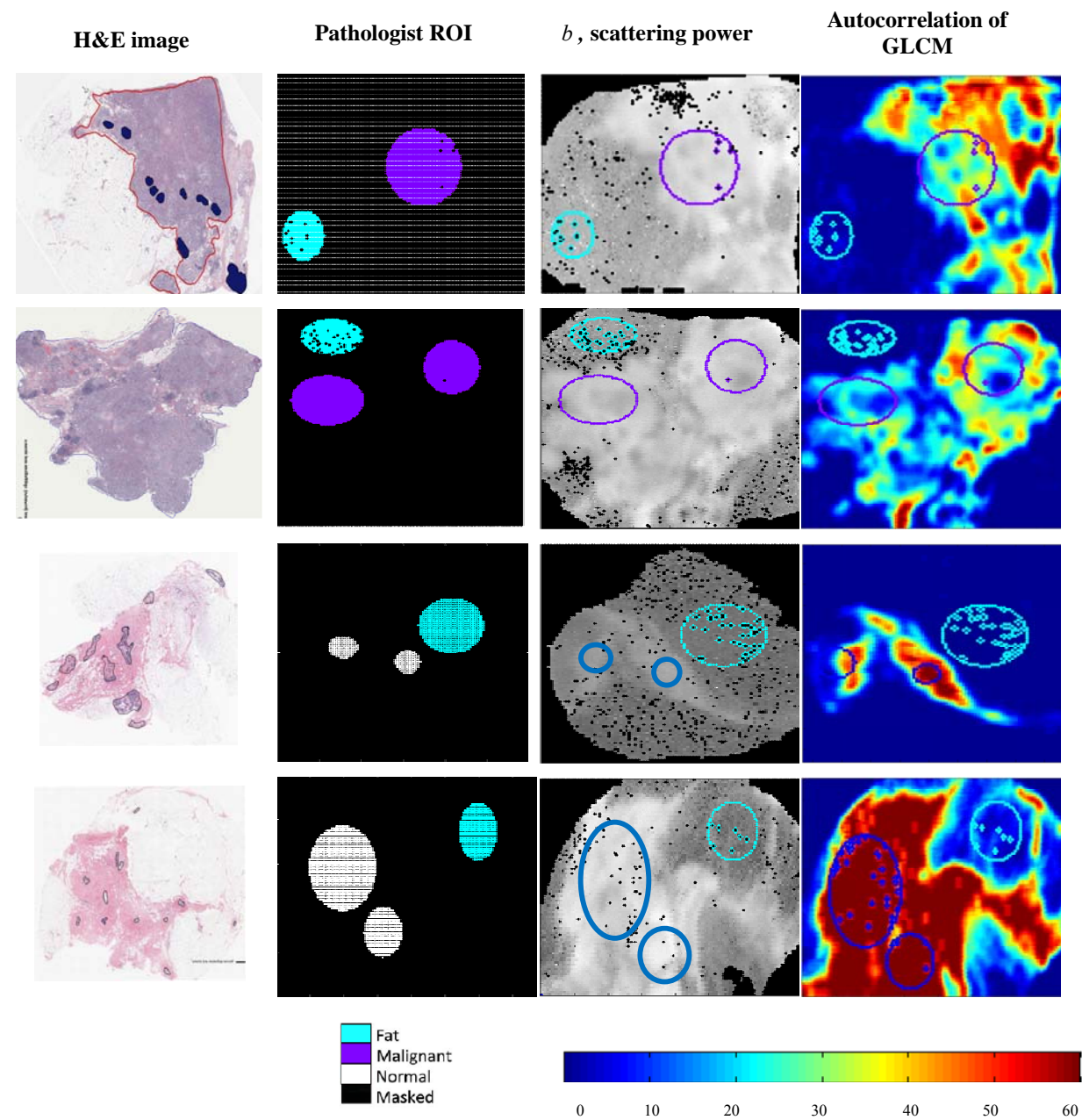

Figure 7. Visual correlation between H&E slides (1st column), pathologist ROI's (2nd column), scattering power map - b (3rd column) and the autocorrelation of GLCM on a 5x5 pixel window (4th column).

4. CONCLUSIONS

A relationship between tissue morphology and scattering power is established by textural analysis of the scattering images. After considering several box plots of different features for initial comparison, autocorrelation of GLMC has shown the best properties to be classified linearly. A sliding window texture analysis based on GLCM property "autocorrelation" has been applied over 48 regions of interest distributed in 29 different samples of fresh breast cancer specimens.

The images obtained from this process seem to be consistent with H&E results and present different values on different regions of interest selected by the pathologist. This permits a linear classifier to achieve sensitivities of 98.74% in adipose tissue, 82.67% in non-malignant tissue and 72.37% in malignant tissue, respectively.

Additionally, as the evaluation of the textural features is not an iterative process, it is perfectly parallelizable. The scattering power images could be previously divided into different computational cores able to evaluate different image regions to speed up the process.

REFERENCES

[1] World Health Organization, 2008: http://www.who.int/en/

[2] Veronesi, U., Cascinelli, N., Mariani, L., Greco, M., Saccozzi, R., Luini, A., Aguilar, M., Marubini, E., "Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer," The New England Journal of Medicine, 347, 1227-1232 (2002).

[3] Fisher, B., Anderson, S., Bryant, J., Margolese, R.G., Deutsch, M., Fisher, E.R., Jeong, J-H., Wolmark, N., "Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer," The New England Journal of Medicine, 347, 1233–1241 (2002).

[4] Laughney, A.M., Krishnaswamy, Rizzo J. Elisabeth, et al. "Scatter Spectroscopic Imaging Distinguishes between Breast Pathologies in Tissues Relevant to Surgical Margin Assessment" Clinical Cancer Research 18(22), 6315 (2012).

[5] Al-Shamlan, H. and El-Zaart, A., "Texture Feature Extraction for Breast Cancer Mammography Images," 2010 International Conference on Image Processing and Computer Vision (IVPCV-10), 69-74 (2010).

[6] Karahaliou, A. N., Boniatis, I. S., Skiadopoulos, S. G., Sakellaropoulos, F. N., Arikidis, N. S., Likaki, E. A., Panayiotakis, G. S. and Costaridou, L. I. , "Breast cancer diagnosis: analyzing texture of tissue surrounding microcalcifications," IEEE Transactions on Information Technology in Biomedicine 12(6), 731-8 (2008).

[7] Kim, K. G., Jong, H. K., Min, B. G. and Bae, K. T., "Segmentation of breast ultrasound lesion boundary using texture-based multi-resolution method," Medical Imaging 2005: Ultrasonic Imaging and Signal Processing, 178-83 (2005).

[8] Yin-Yin Liao, Po-Hsiang Tsui, Chia-Hui Li, King-Jen Chang, Wen-Hung Kuo, Chien-Cheng Chang and Chih-Kuang Yeh , "Classification of scattering media within benign and malignant breast tumors based on ultrasound texture-feature-based and Nakagami-parameter images," Med.Phys. 38(4), 2198-207 (2011).

[9] Doyle, S., Agner, S., Madabhushi, A., Feldman, M. and Tomaszewski, J., "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," 2008 IEEE International Symposium on Biomedical Imaging: From Macro to Nano (ISBI '08), 496-9 (2008).

[10] Lu, R. and Peng, Y. , "Multispectral scattering measures fruit texture," Laser Focus World 41(4), 99-103 (2005).

[11] Cohen, F. S., Taslidere, E. and Hari, D. S., "Tissue characterization and detection of dysplasia using scattered light," 2006 3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano, 590-3 (2006).

[12] Krishnaswamy, V., Hoopes, P.J., Samkoe, K.S., O'Hara, J.A., Hasan T., Pogue, B.W., "Quantitative imaging of scattering changes associated with epithelial proliferation, necrosis, and fibrosis in tumors using microsampling reflectance spectroscopy," Journal of Biomedical Optics 14(1), 014004 (2009).

[13] Laughney, A.M., Krishnaswamy, V., Garcia-Allende, P.B., Conde, O.M., Wells, W.A., Paulsen, K.D., Pogue, B.W., "Automated classification of breast pathology using local measures of broadband reflectance," Journal of Biomedical Optics 15(6), 066019 (2010).

[14] Jacques, S.L., Prahl, S., Oregon Medical Laser Center. 2010.

[15] Garcia-Allende, P.B., Krishnaswamy, V., Hoopes, P.J., Samkoe, K.S., Conde, O.M., Pogue, B.W., "Automated identification of tumor microscopic morphology based on macroscopically measured scatter signatures," Journal of Biomedical Optics 14(3), 034034 (2009).

[16] Haralick, R. M., Shanmugam, K. and Dinstein, I. , "Textural features for image classification," IEEE Transactions on Systems, Man and Cybernetics smc 3(6), 610-621 (1973).

[17] Soh, L. and Tsatsoulis, C. , "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," IEEE Trans.Geosci.Remote Sens. 37(2), 780-795 (1999).