TEXTURAL ANALYSIS OF OPTICAL SCATTERING FOR IDENTIFICATION OF CANCER IN BREAST SURGICAL SPECIMENS

Alma Eguizabal [(1)], Ashley M. Laughney [(2)], Pilar Beatriz García-Allende [(1,3)], Venkataramanan Krishnaswamy[(2)], Wendy A. Wells [(4)], Keith D. Paulsen[(2)], Brian W. Pogue[(2)], Jose M. Lopez-Higuera[(1)], Olga M. Conde[(1)]

[(1)]Grupo Ingeniería Fotónica, Univ. de Cantabria (Spain); [(2)] Thayer School of Engineering at Dartmouth (United States); [(3)]Helmholtz Zentrum München GmbH (Germany); [(4)] Dartmouth Hitchcock Medical Ctr. (United States)

ABSTRACT

Textural analysis of tissue scattering images is proposed for healthy versus tumor discrimination. Scattering center density varies from normal to tumor tissues and this variation is translated intodifferent textures in the scattering power map. Adipose tissue shows low autocorrelation values while tumor tissues present higher entropies than normal tissue. Consequently, a combination of autocorrelation and entropy values allows ready tissue discrimination by a supervised linear classifier. The proposed approach has been validated over a set of 29 breast tissue samples achieving a sensitivity of 73.59% and specificity of 82.40%.

Index Terms— breast tumor delineation; textural analysis; optical scattering; autocorrelation; entropy.

1. INTRODUCTION

Breast cancer is the most common female cancer worldwide. It is estimated that one over eight women in the world will be diagnosed with breast cancer [1]. Nevertheless, cancer pathology presents higher survival rate if it is detected at an early stage.

Breast-Conserving Therapy (BCT), which includes local excision and radiation treatment to the breast, has been the standard of care for early invasive breast cancers and has been demonstrated to be equally safe and effective as mastectomy for most patients when surgical margins are clear of residual disease [2]. However, surgical margin assessment lacks standardization, which is a high risk to patient's long-term prognosis.

A scanning spectroscopy platform has been designed to selectively sample the scattering response of breast surgical specimens; with the ultimate goal of developing a diagnostic adjunct to surgical procedures [3]. Spectrally, light scattering is relatively featureless; however, spatially its distribution reflects the natural heterogeneity observed in tissue morphology. The scattering spectrum provides information about the size and number density of cellular components and structural matrix elements [4]; which spatially transform with malignancy.

Several studies have applied texture analysis in breast tissue analysis and characterization [5-7]. Most work on computer-aided diagnosis (CAD) for breast cancer has focused on the detection of tumor masses in mammograms [5][7] or the diagnosis of cancer in microscopic pathology, the diagnostic gold standard [6]. Standard histopathologic processing requires fixation of the tissue specimen and interpretation by a pathologist, which takes several days and is consequently performed after surgery. The work here aims to automate detection of malignancy in fresh tissue specimens at the time of primary surgery to allow for a more complete resection. The optical assembly samples weakly scattered photons which are sensitive to the ultra-structural changes associated with disease; and its scanning design allows for assessment of their spatial distribution.

This paper will focus on how conventional textural analysis of the slope of the scattering spectrum, could be finally applied to evaluate the probability of malignancy with a simple, fast and linear classification to obtain an accurate bound. Regions of Interest (ROI) contrasted by a pathologist have been used to validate the classification.

2. OPTICAL DATA

The scatter imaging system consisted of a weakly confocal spectroscopic system with illumination and detection spot sizes approaching one mean scattering length (typically 100 μm for tissue [8]), and a raster-scanning platform built using linear translation stages. The instrument operates in the 510 nm – 785 nm spectral waveband with a broadband fibre-coupled tungsten-halogen light source. An optical-fibre coupled to a CCD-based spectrometer was used for confocal spectroscopic detection, with a spectral resolution of approximately 1 nm. All spectral measurements were referenced to a Spectralon based reflectance standard, which removes instrumental spectral response from sample measurements. A schematic and a more detailed description of the system can be found in this previous paper [9] where the same set-up and data was used with a supervised classifier.

An empirical approximation to Mie theory (power law dependence on wavelength) was used to describe the relative reflectance due to light scattering and an exponential factor was included to correct for local pooling of hemoglobin, a dominant tissue absorber in the visible waveband [10].

$$I_R = A\lambda^{-b} e^{(-k \cdot c \cdot (d \cdot HbO_2(\lambda) + (1-d) Hb(\lambda)))} \quad (1)$$

A is the scattered amplitude, *b* the scattering power, and the exponential correction factor depends on the path length (*k)* a constant proportional to the concentration of whole blood (*c)* and the oxygen saturation fraction (*d)*. The illumination-detection geometry combined with the typical values of absoroption encountered in tissue, limits detected photons to those primarily influenced by scatter (pathlengths are too short for probably absorption). Scattering parameters related to tissue morphology are *A* and *b*. During the fitting process some pixels had a poor data fit to the empirical model and they become masked.

This study is focused on the analysis of 29 tissue samples, with 48 different Regions of Interest (ROIs) diagnosed by a pathologist (WAW) and corresponding to 7 different pathologies. All these categories, summarized in Table I, were classified more generally as non-malignant (normal or benign), malignant and adipose.

TABLE I. Regions of interest diagnosed by the pathologist.

Tissue type	No of ROI	No of pixels
Non-Malignant	***25***	***36979***
Malignant	***14***	***23220***
Adipose	***9***	***7021***
Total ROI	***48***	***67220***

3. TEXTURAL ANALYSIS

In general terms, image texture can be defined [10] as the spatial variation in pixel intensities, in this case, of the scattering parameter power values. Optically, fatty tissues have large scatterers leading to low scattering powers, wherea glandular or fibrous tissues have more cellular and collagen-based structures that lead to high scattering powers [11]. One immediate application of image texture is therefore image region segmentation, i.e. tissue diagnosis, using texture properties.

The scattering power can be interpreted as a gray-level variation that differs from one tissue to another, as depicted in Figure 1. The definition of texture properties requires the preliminar computation of the so-called co-ocurrence matrix. This matrix represents a two-dimensional histogram to evaluate the image properties related to second-order statistics of gray values along the texture.

$$C(k,l,;d,\phi) = \sum\sum \partial(k - g(i,j)) \partial(l - g(i + d\cos\phi, j + d\sin\phi)) \quad (2)$$

For the co-ocurrence matrix expression (2), $g(i,j)$ is the grey value of pixel (i,j) considering 8 gray level values to map the whole range of scattering power values. The computation angle, ϕ , and the distance, d , between pixels need to be specified. Studies performed on the scattering power maps have shown that there is no a clear orientation in the scattering power textures and, on the other hand, a distance of one pixel has been used. Once the co-ocurrence matrix is computed, and normalized by dividing all its elements by the total number of pairs considered, the joint probability density $p(i,j)$ is obtained. Several metrics can then be derived from it [10-11]: energy, correlation, entropy, homogeneity, autocorrelation, etc.

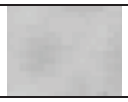
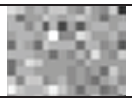
Scattering Power Map			
Tissue Diagnosis	Non-Malignant	Malignant	Adipose

Figure 1.Examples of scattering power maps of different tissue types: non malignant, malignant and adipose.

3.1. Entropy

This feature is expected to arise higher values for more complex images [10]:

$$f_2 = -\sum_i \sum_j p(i,j) \log(p(i,j)) \quad (3)$$

Malignant tissues are typically characterized by increased epithelial cell density, non-uniform nuclear size, and degredation of the extracellular matrix. It is intuitive to expect their texture to have more entropy because of this increased disorder. As seen in Figure 2, adipose tissue also presents high entropy values. This could introduce an extremely high probability of false alarms. This is the reason why we decided to extract another feature to detect and mark all adipose pixels and reduce the false alarm rate.

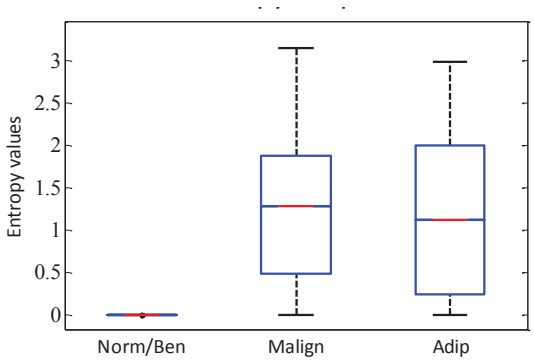


Figure 2. Boxplots of entropy values corresponding to diagnosed pixels of different tissue categories.

3.2. Autocorrelation

This feature provides a measure of gray-tone linear-dependencies in the image [11]:

$$f_1 = \sum_i \sum_j (i\,j)\, p(i,j) \qquad (4)$$

Adipose tissue tends to behave with an irregularity along the spatial domain of the sample, which finally characterize it with a lower autocorrelation value.

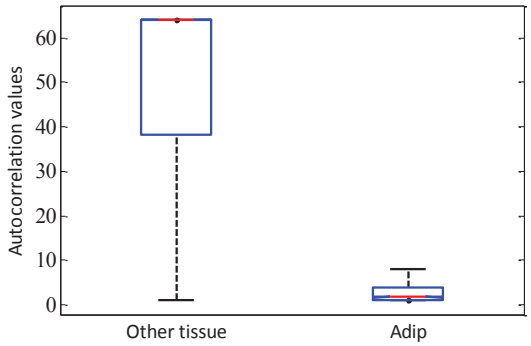


Figure 3. Boxplots of autocorrelation values that reveals how adipose tissue differs from other tissue categories.

3.3. Spatial domain for texture analysis

Texture analysis requires the definition of a spatial domain around the pixel to be able to compute the co-ocurrence matrix and their associated metrics. A study about the optimum spatial size performed in [9] concluded an optimized window dimension of 5 pixels. This size corresponds to a 500 μm side square window. To this end, a sliding window of 5x5 pixels is centered on each pixel with the purpose of obtaining the texture of enough tissue to be correctly diagnosed, but not too much so it is safely bounded. The center pixel is then assigned with the computed texture features to create a new image of the same size.

4. MALIGNANCY IDENTIFICATION

Malignancy identification is accomplished by two simple and fast linear classifiers: the first applies a threshold to segment adipose tissues, resulting in a lower false positive rate; the second detects malignant pathologies in the imaged field.

One generic supervised threshold computation for malignancy identification is the mean value of the lower quartile of normal tissue and the upper quartile of tumor tissue entropies. Adipose tissue has to be previously marked with a similar procedure. Figure 4 shows a block diagram of the whole process.

5. RESULTS AND DISCUSSION

A leave-one-out approach was used on every ROI of the 29 samples to evaluate the classification performance. Figure 5 shows how both thresholds, adipose identification based on autocorrelation threshold and malignancy identification based on entropy threshold, remain stable along the samples. Considering the threshold levels on the leave-one-out process, a mean sensitivity of 73.59% and specificity of 82.40% have been achived, with best pair of values obtained as 98.94% sensitivity with 98.90% specificity on one of the samples.

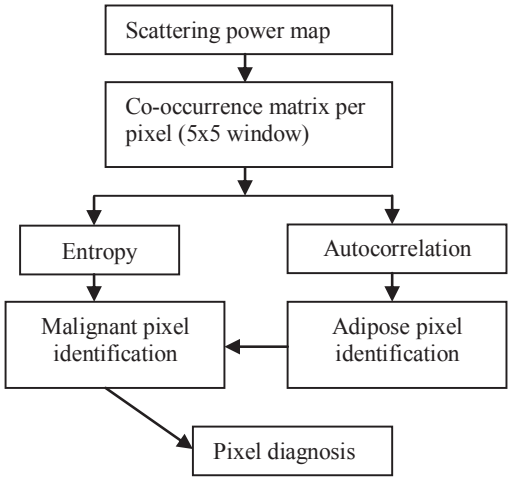


Figure 4. Malignancy identification process.

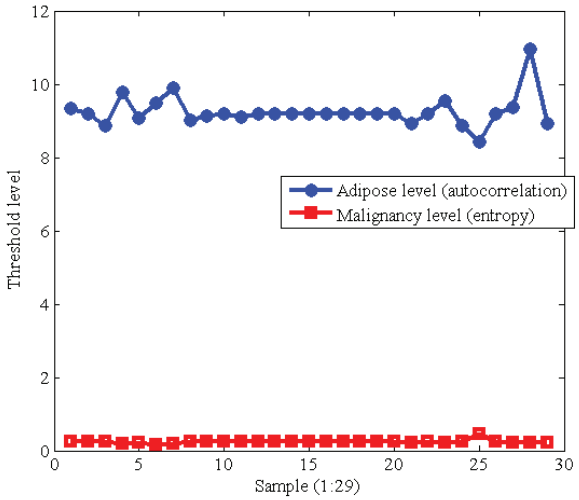


Figure 5. Stability of the decision thresholds.

Different levels of the threshold were performed to study its effect on the classifier; however, this is slightly biased because the number of pixels of each diagnosis was not the same (Table I). ROC curves of malignancy classification were extracted from different probabilities of adipose detections. These are conditionated to two decisions so sometimes are incompleted. As sensitivity of adipose tissue increases, the specifity of tumor improves. But if

sensivity of adipose is too high, detection of malignancy presents limitations.

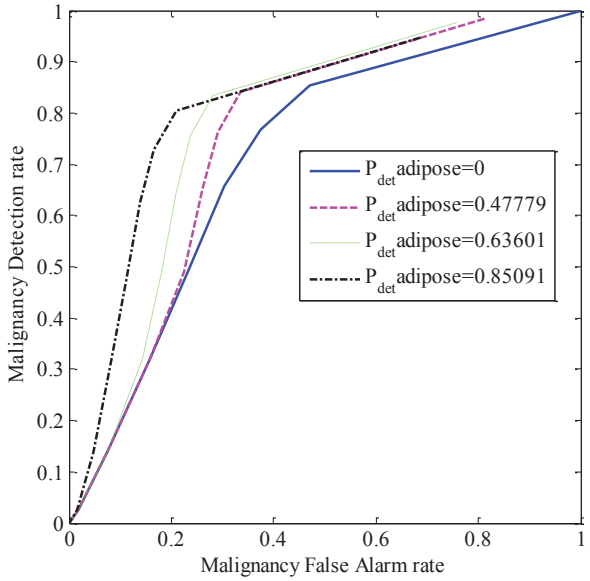


Figure 6.ROC curve of malignancy identification.

6. CONCLUSIONS

Breast tumor scattering maps have been characterized with entropy and autocorrelation textural features in order to identify malignancy in scattering images of breast surgical specimens. Only linear classification, based on threshold level evaluations, has been used. ROC estimation suggests that sensitivities and specificities higher that 80% could be obtained with an optimal threshold. The selection of the optimal threshold is based on a preliminary supervision that has exhibited independency on interpatient variability. Malignancy identification is rapidly attained since the decision is just based in a linear evaluation of a threshold. Furthermore, as textural feature computation is not dependent on previous results, i.e. it is not an iterative process, malignancy identification is easily parallelizable. The scattering power images could be previously divided into different computational cores able to evaluate different image sections and different image features.

The validation has only considered pixels on ROIs so an accurancy on tumor margins can not be demonstrated. However visual correlation between H&E results and texture maps may suggest that margin delineation would be the next implementation of the system.

ACKNOLEDGEMENTS

This work has been supported by CYCIT projects DA2TOI (FIS2010-19860) and TFS (TEC2010-20224-C02-02), funded by the Spanish Government.

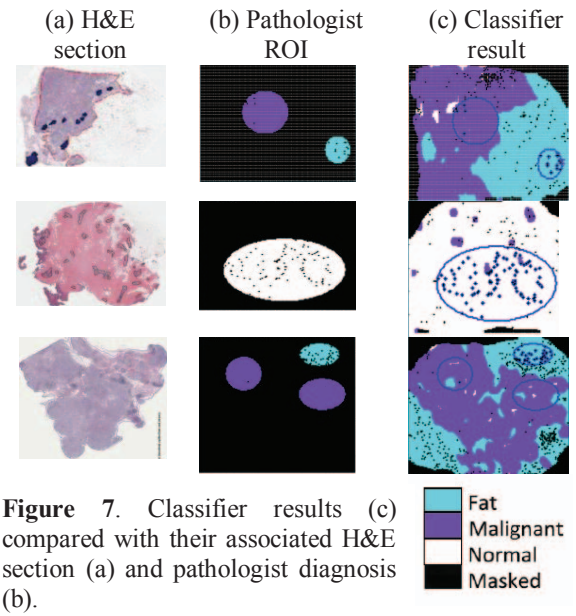


Figure 7. Classifier results (c) compared with their associated H&E section (a) and pathologist diagnosis (b).

REFERENCES

[1] AECC, Asociación Española Contra el Cancer http://www.aecc.es.

[2] F. Fitzal, O. Riedl and R. Jakesz, "Recent developments in breast conserving surgery for breast cancer patients" *Langenbeck's Archives of Surgery,* 394(4), 591-609 (2009).

[3] V.Krishnaswamy et al., "Quantitative imaging of scattering changes associated with epithelial proliferation, necrosis, and fibrosis in tumors using microsampling reflectance spectroscopy ", *Jour. of Biomed Optics*, 14(1), 014004, 2009.

[4] S.Srinivasan *et al.*, "Interpreting hemoglobin and water concentration, oxygen saturation, and scattering measured in vivo by near-infrared breast tomography," *PNAS*, 100(21), 12349-12354, 2003.

[5] M.E.Mavroforakis *et al.*, "Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers," *Artif. Intell. in Medicine,* 37(2), 145-162, 2006.

[6] S.Doyle *et al.*, "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features," *5th IEEE ISBI: From Nano to Macro (ISBI 2008)*, 496-499, 2008.

[7] A.Oliver *et al.*, "A novel breast tissue density classification methodology," *IEEE Trans.Inf.Technol.Biomed.*, 12(1), 55-65, 2008.

[8] B.W.Pogue, G.C.Burke, "Fiber-optic bundle design for quantitative fluorescence measurement from tissue," *Applied Optics,* 43, 3048-3054, 2004.

[9] A.M.Laughney *et al.,* "Automated classification of breast pathology using local measures of broadband reflectance", *Jour. of Biomedical Optics*, 15 (6), 066019-066019-12, 2010.

[10] R.M.Haralick, K.Shanmugam, Its'Hak Dinstein.,"Textural Features for Image Classification," *IEEE Trans. On Systems, Man and Cybernetics,* 3 (6), 610-621, 1973

[11] L.K.Soh, C.Tsatsoulis, "Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices," *IEEE Trans. on Geoscience and Remote Sensing,* 37 (2), 780-795, 1999.