

Automated interpretation of scatter signatures aimed at tissue morphology identification

P.B. Garcia-Allende^a, V. Krishnaswamy^b, K.S. Samkoe^b, P.J. Hoopes^{b,c}, B.W. Pogue^{b,c}, O.M. Conde^{*a}, J.M. Lopez-Higuera^a

^aPhotonics Engineering Group, University of Cantabria, Avda. Los Castros S/N, 39005 Santander, Spain;

^bThayer School of Engineering, Dartmouth College, Hanover, NH, USA 03755;

^cDepartment of Surgery, Dartmouth Medical School, Hanover, NH, USA 03755

ABSTRACT

An automated algorithm and methodology is presented to pathologically classify the scattering changes encountered in the raster scanning of normal and tumor pancreatic tissues using microsampling reflectance spectroscopy. A quasi-confocal reflectance imaging system was used to directly measure the tissue scatter reflectance *in situ*, and the spectrum was used to identify the scattering power, amplitude and total wavelength-integrated intensity. Pancreatic tumor and normal samples were characterized using the instrument and subtle changes in the scatter signal were encountered within regions of each sample. Discrimination between normal vs. tumor tissue was readily performed using an Artificial Neural Network (ANN) classifier algorithm. A similar approach has worked also for regions of tumor morphology when statistical pre-processing of the scattering parameters was included to create additional data features. This automated interpretation methodology can provide a tool for guiding surgical resection in areas where microscopy imaging do not reach enough contrast to assist the surgeon.

Keywords: tumor, necrosis, confocal reflectance imaging, feature extraction, automatic classification, Artificial Neural Networks

1. INTRODUCTION

Scatter visualization of tissue ultra-structure *in situ* could provide a unique tool for guiding surgical resection. In a previous study [1], scattering changes associated with epithelial proliferation, necrosis and fibrosis in tumors were estimated in a quantitative manner using highly localized measurements of reflectance. Since encountered changes were subtle, an automated methodology is required to establish a correlation with tissue morphologies before proceeding to clinical studies. A great variety of methods are suitable to perform this interpretation into what the signal means relative to the pathology such as the *k*-nearest neighbors (kNN) algorithm [2] that provided a good classification accuracy inside some predefined regions of interests [3]. However, to assess how useful this algorithm is, a comparison with other types of data discriminators such as Principal Component Analysis (PCA) [4] or Linear Discriminant Analysis [5] has still to be presented. Apart from the small changes in scattering, data is multi-parametric, and thus Artificial Neural Networks (ANNs) [6] could prove successful in tissue sub-type classification due to their ability to handle non-linearity. The validity of this assumption not only for tumor versus normal tissue discrimination, but also for tumor region segmentation, is demonstrated here.

2. TISSUE SCATTER IMAGING

2.1 Scatter imaging system and parameter fitting

The scatter imaging system consisted of a weakly confocal spectroscopic system having illumination and detection spot sizes smaller than one mean scattering length (typically 100 μm for tissue [6]), 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 the sample measurements and allows direct comparisons of the extracted scatter parameters across different samples. A schematic and a more detailed description of the system can be found in this previous paper [1].

The acquired spectral reflectance is fitted by a empirical relationship, which accounts for scatter versus wavelength, and then exponential attenuation due to absorption

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

where A is the scattered amplitude, b the scattering power, k is the path length, c is a constant proportional to the concentration of whole blood and d is the oxygen saturation fraction. Absorption contributions from other chromophores, other than hemoglobin, are assumed to be negligible in the waveband of interest (from 510 to 785 nm as mentioned before) and k is assumed to be wavelength independent constant. The extinction spectra of oxygenated and deoxygenated hemoglobin, $HbO_2(\lambda)$ and $Hb(\lambda)$, were obtained from the Oregon Medical Laser Center database. Scattering parameters of interest for tissue morphology identification are A and b , along with the so-called average scattered irradiance, I_{avg} , which was calculated by integrating I_R over wavelengths beyond 610 nm to avoid the hemoglobin absorption peaks.

2.2 Pancreatic tumors

Human pancreatic tumor cells ASPC-1 were grown and injected subcutaneously in the flank region of male mice [1]. Seven weeks after, tumors were harvested and imaged using the raster-scanning reflectance imager described previously. After the measurement, the top full view slide from each sample was used for pathologic analysis and their H&E slides were examined by a veterinary pathologist who identified several regions-of-interest corresponding to the observed tissue types (normal or tumor) and tumor regions. These were classified under three major groups, namely epithelium, fibrosis and necrosis, with different constituent subgroups: two distinct types of epithelial cells according to the exhibited nucleus to cytoplasm ratio were considered (low and high proliferation index) and regions exhibiting fibrosis were classified into early, intermediate and mature fibrosis subgroups. Figure 1 depicts one of the analyzed pancreas tumor samples, where five regions-of-interest are shown overlaid on the scattering parameter images (scattered amplitude A , scattering power b and average scattered irradiance I_{avg} whose color scale bars are shown on the right of each corresponding image). Example histology images of each region have shown [1] how tumor cells found in Region 1 have less cellular density (LPI) than the HPI tumor cells found in Region 2. Region 3 exhibits necrosis and Regions 4 and 5 early and intermediate fibrosis, respectively. Pixels in black correspond to locations where the scatter data could not be reliably measured and they are tagged from now on as *masked pixels*.

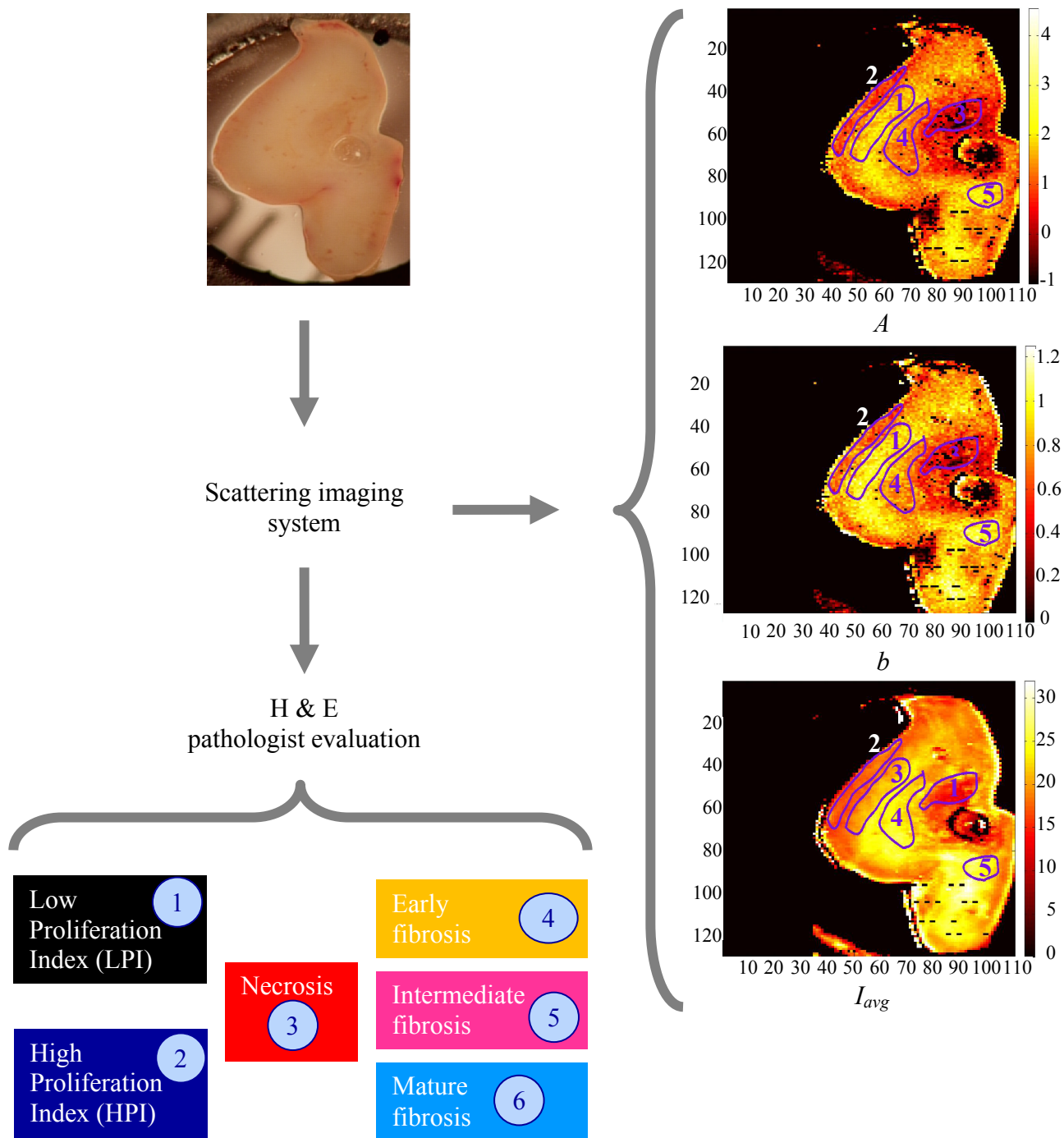


Fig. 1. Scattered amplitude (*A*), scattering power (*b*) and average scattered irradiance (*I_{avg}*) images of a pancreas tumor sample showing five pathology based regions-of-interest overlaid on it.

2.3 Classification methods

An ANN was properly trained to be able to distinguish first normal tissue and tumor morphologies. The input data set consisted of the three mentioned scattering parameters (the scattered amplitude, A , the scattering power, b , and the average scattered irradiance, I_{avg}) of different pixel localizations within the predefined regions-of-interest of three normal and five tumor samples. The three scattering parameters of interest were fitted as explained above. A multilayer feed-forward network and back-propagation learning algorithm were employed. The topology of the net consisted of an input layer with 10 neurons, and an output layer with only 1 neuron. Neurons in both layers have a log-sigmoid transfer function. The activation of the neuron in the output layer, which means that the output value is close to 1, indicates the presence of a tumor pixel localization.

An alike attempt to segment the distinct tumor sub-types (low proliferation index epithelium, high proliferation index epithelium, necrosis, early fibrosis, intermediate fibrosis and mature fibrosis) was accomplished employing a data set that included the scattering parameters (A, b, I_{avg}) of pixel localizations within all possible tumor regions. However, not enough discrimination performance was achieved and only necrotic regions were correctly identified. In [3] the specificity in tumor sub-type discrimination was remarkably improved by the inclusion of a pre-processing procedure to generate additional features. These were based upon statistical estimates (mean, standard deviation, skewness and kurtosis) [8] of the scattering parameters between neighboring pixels within a window of interest. The latter is a moving window, which was scanned around the image to generate statistical values for each pixel, and whose size was adjusted to assure that statistics are predominantly calculated for pixel localizations within the same tissue sub-type [3]. As the tumor segmentation capability of ANNs in the (A, b, I_{avg}) space was weak, statistics were also here concatenated with the fitted scattering parameters to form a 15-dimensional feature space. In addition, the complexity of the network topology was increased to achieve reliable classification results (two hidden layers, with 50 neurons each, were added and the output layer consisted of six neurons corresponding to each tumor sub-type to be discriminated). Figure 2 illustrates a comparison between the topologies of the networks employed for the discrimination between normal and tumor tissue and the segmentation of pathologically distinct tumor regions, respectively. As mentioned before only, the scattering parameters are input to the first normal-vs-tumor network. Input data of the second tumor-sub-type network includes the statistics, where \bar{p} and σ_p are the mean and variance of the scattering parameter p , and S_p and K_p stand for its skewness and kurtosis moments. The ANN results that match each tumor subtype and the transfer function of the neurons are also depicted in Figure 2.

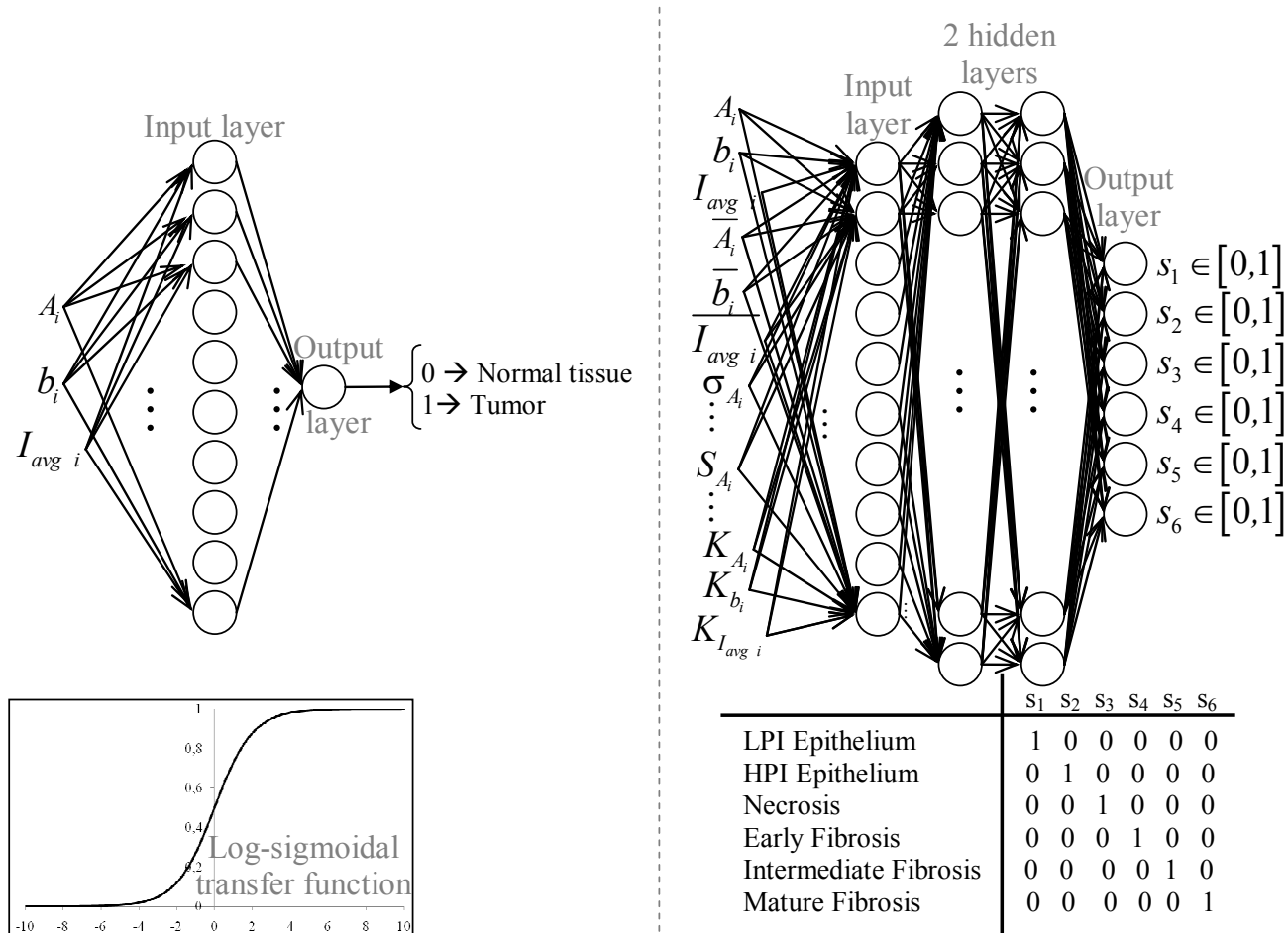


Fig. 2. ANN topologies and outputs for tumor vs. normal tissue (on the left) and tumor sub-type (on the right) discrimination. Inset: Transfer function of the neurons.

3. RESULTS

3.1 Tissue morphology identification

Figure 3 depicts the grouped scatter plot for normal and tumor tissue pixel localizations in the scattering parameter feature space (A - b - I_{avg} space) showing that scattering parameters of normal pixels are well grouped, whereas there is a remarkable spreading of tumor pixels. A total of 16599 pixel localizations within the regions-of-interest of three normal samples and 3660 pixels within the regions-of-interest of five tumor samples were considered. This data set was divided into 3 non-overlapping sets containing 6653 data points each (5533 of normal tissue and 1120 of tumor). Two of these sets were employed as the training data set, i.e. it is input to the neural network which iteratively adjust its parameters (weights and biases) in order to produce an output that matches the desired result ('0' for normal tissue pixels and '1' for tumor). The third set, the so-called validation set, is employed to measure the error of the network in tissue type identification. Only one pixel location out of 6653 test data points was misclassified. A specificity of the designed network for tissue type identification of 99,98% is therefore demonstrated. Finally, Figure 4 depicts the discrimination results for all pixel locations in normal (left column) and tumor (right column) samples. The pixels tagged as *masked* correspond to pixels where the scatter data could not be reliably measured.

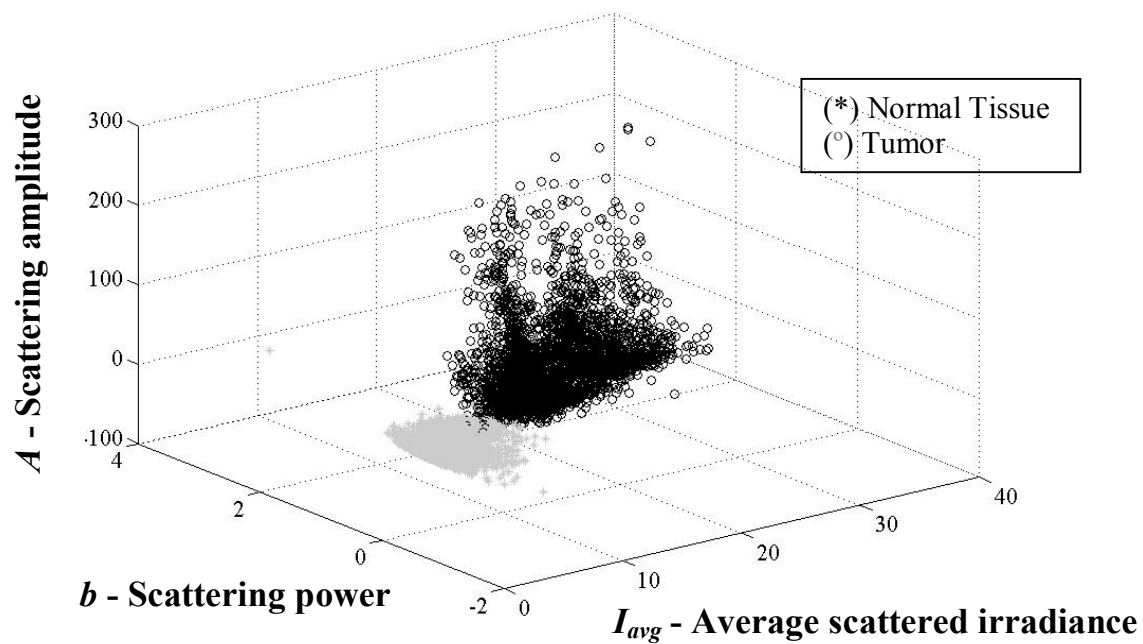


Fig. 3. 3D map representing the training normal and tumor pixel localizations in the three input data parameters, A - b - I_{avg} , space.

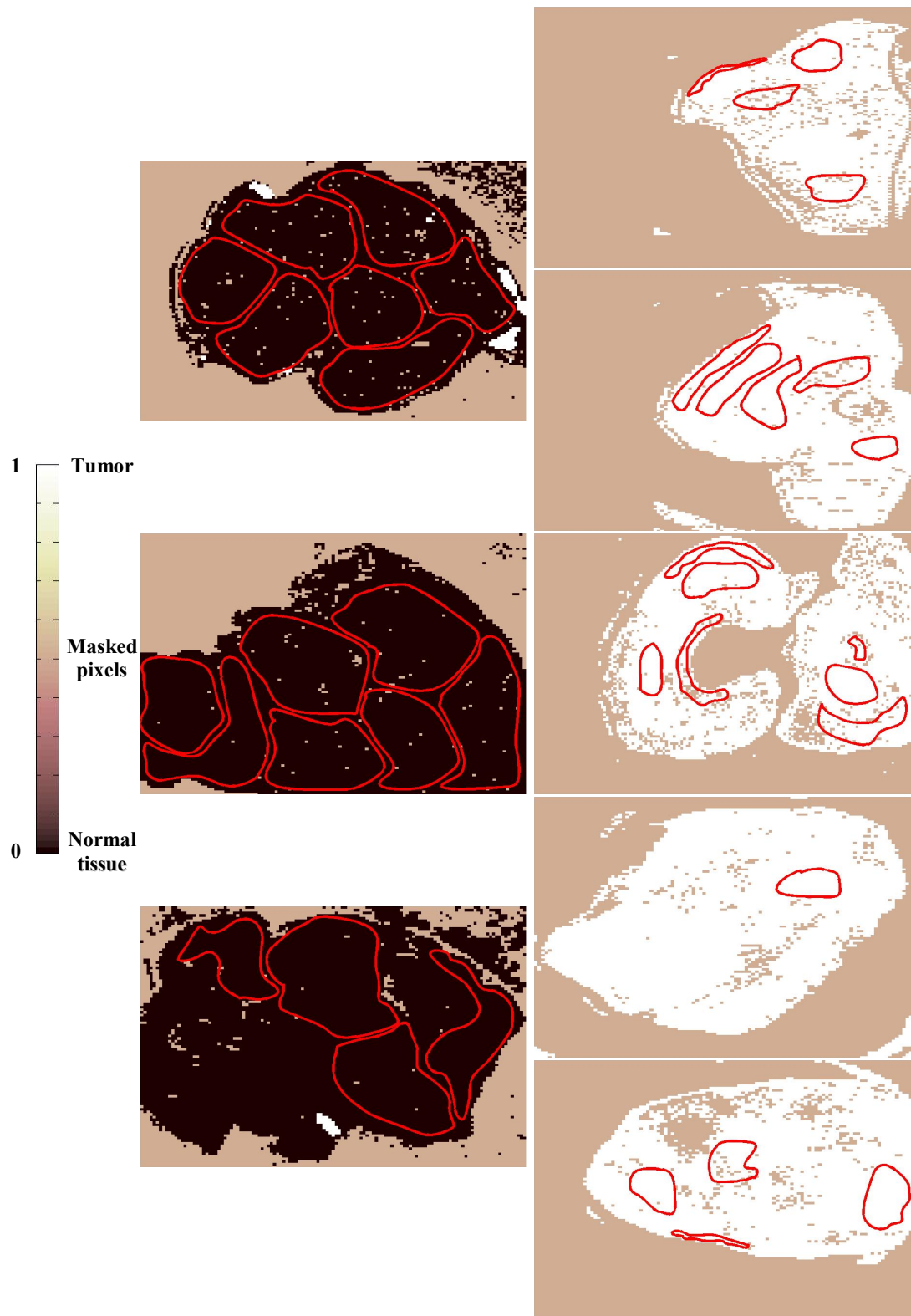


Fig. 4. Pseudo-color diagnosis images using the three measured scatter parameters (normal samples on the left and tumor samples on the right).

3.2 Tumor sub-type delineation

The 3D map representing tumor pixels in the A - b - I_{avg} space is presented in Figure 5. As shown, scattering parameters are rather grouped only in necrotic locations. Therefore, the training of the neural network only converged when discriminating necrosis against the rest tumor types. Thus, the segmentation of the tumor regions could be tackled by bridging of results from a combinatorial set of one-against-all classifiers [7]. To directly deal with the multiclass classification problem the statistics-based feature extraction procedure introduced in *Classification Methods* was implemented. A schematic and a more detailed description of the latter can be found in [3]. The larger the dimension of the vicinity region employed in statistics calculation, they are more accurate. However, this only happens when pixel localizations within the window are mostly pathologically identical, i.e. inside the same tissue sub-type. This was assured by computing the histogram of the width and length of the regions of interest identified by the pathologist. As shown in [3] it was not advisable to employ window sizes larger than 15 since they had low occurrence probabilities. A square vicinity region of 12 x 12 pixels was selected in the proposed approach.

The identification of the regions-of-interest performed by the veterinary pathologist and the automated segmentation based upon ANNs for tumor samples presented in Figure 4 are depicted in Figure 6. As shown, a strong correlation is achieved with the expert-based classification within the predefined regions-of-interest.

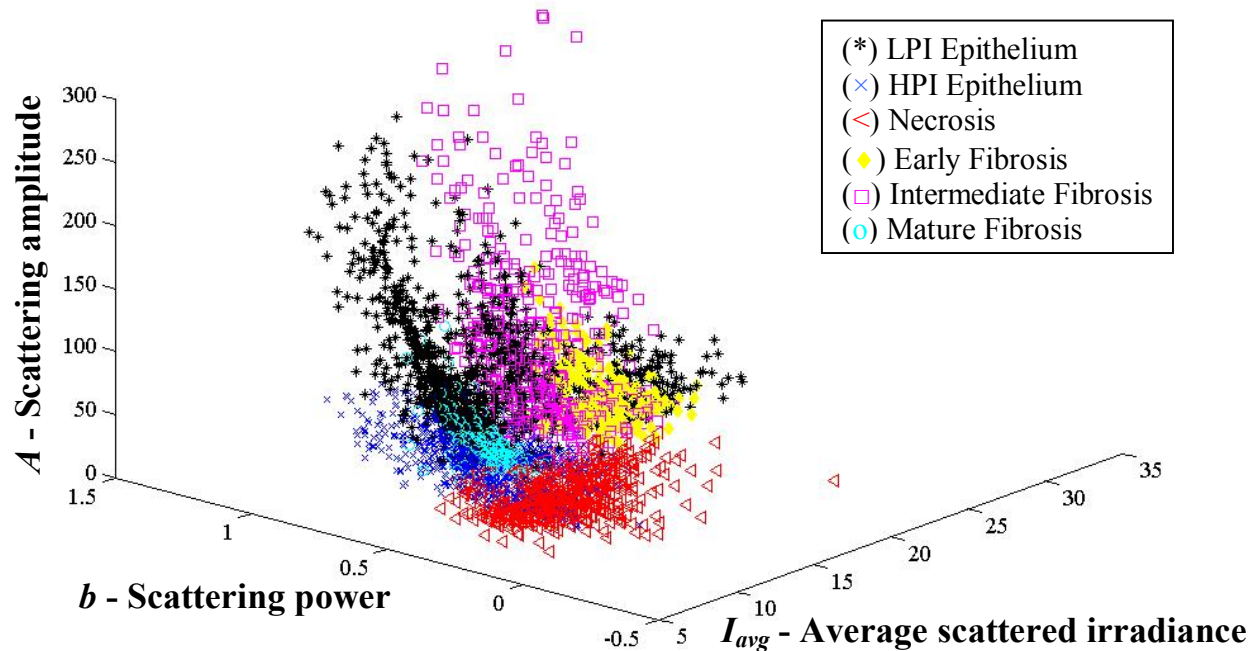


Fig. 5. Grouped scatter plot for all tumor pixel localizations in the A - b - I_{avg} space.

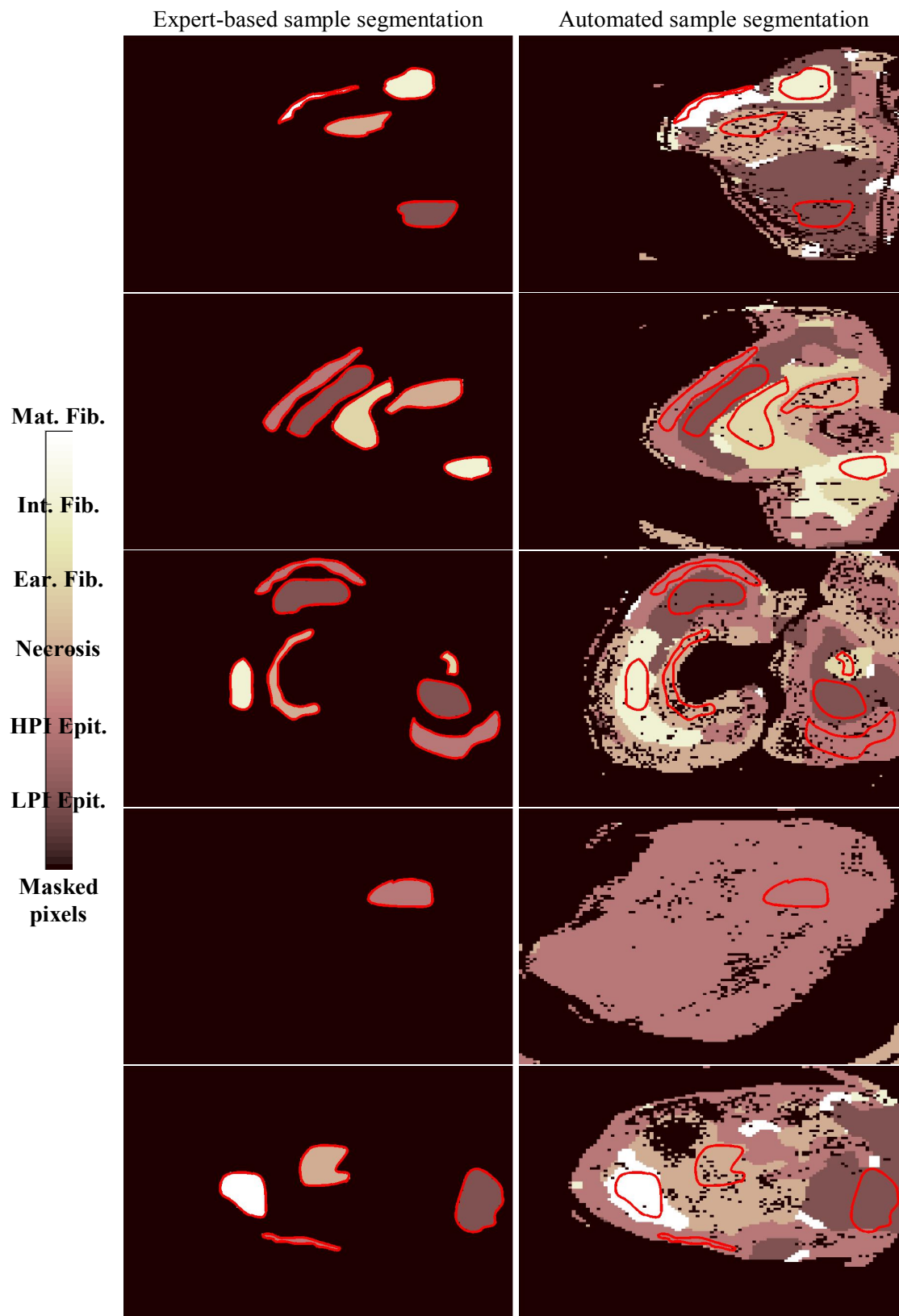


Fig. 6. Qualitative comparison between expert- and automated-based segmentation of the tumor samples.

4. CONCLUSIONS

An automated methodology based upon ANN classifiers to quantify morphology-associated scatter changes in tissue has been designed. At first, the capability for quantitative tissue type identification, i.e. tumor versus normal tissue discrimination, was demonstrated. Tumor sub-type segmentation based on ANN with the initial scattering parameters (scattering amplitude, scattering power and average scattered irradiance) showed a weak performance. The inclusion of the high-order statistical parameters of each fitted scattering parameter was required to achieve a good correlation between expert-based and automated region of interest identification and an adequate specificity in tumor delineation. Quantitative comparisons with the previously developed algorithm [3] are in progress. Additionally, it should be mentioned that, although the feasibility of the proposed technique has been demonstrated, further studies considering different tissue types are required before it could gain clinical adoption. Specifically, further studies regarding breast tissue are currently being conducted.

ACKNOWLEDGMENTS

P. Beatriz Garcia Allende is grateful for financial support from the Spanish Government's Ministry of Science and Technology through the project TEC'2007-67987-C02-01 allowing her to spend three months in the Thayer School of Engineering at Dartmouth College. Funding at Dartmouth for this work came from NIH research grants PO1CA84203 and PO1CA80139.

REFERENCES

- [1] Krishnaswamy, V., Hoopes, P.J., Samkoe, K.S., O'Hara, J.A., Hasan, T., and 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).
- [2] Fukunaga, K., [Introduction to statistical pattern recognition], Academic Press, 2nd ed. (1990).
- [3] Garcia-Allende, P.B., Krishnaswamy, V., Hoopes, P.J., Samkoe, K.S., Conde, O.M. and Pogue, B.W., "Automated identification of tumor microscopic morphology based upon macroscopically measured scatter signatures," *Journal of Biomedical Optics* 14(3), *In press*.
- [4] Workman Jr, J. and Springsteen, A.W., [Applied spectroscopy: a compact reference for practitioners], Academic Press Limited, 1st ed. (1996).
- [5] Michie, D. and Spiegelhalter, D.J. and Taylor, C.C., [Machine Learning, Neural and Statistical Classification], Prentice Hall, 1st ed (1994).
- [6] Chen, C.H., [Fuzzy logic and neural network handbook], McGraw Hill (1996),
- [7] Pogue, B.W. and Burke, G.C., "Fiber-optic bundle design for quantitative fluorescence measurement from tissue," *Applied Optics* 43, 3048-54 (2004).
- [8] Oliver, A., Freixenet, J., Martí, R., Pont, J., Pérez, E., Denton, E.R.E. and Zwigelaar, R., "A novel breast tissue density classification methodology," *IEEE Transactions on Information Technology in Biomedicine* 12 (1), 55-65 (2008).