Enhanced tumor contrast during breast lumpectomy provided by independent component analysis of localized reflectance measures

Alma Eguizabal[a], Ashley M. Laughney[b], Pilar Beatriz Garcia Allende[c], Venkataramanan Krishnaswamy[b], Wendy A. Wells[d], Keith D. Paulsen[b], Brian W. Pogue[b], Jose M. Lopez-Higuera[a], Olga M. Conde[a]
[a]Univ. de Cantabria (Spain);
[b]Thayer School of Engineering at Dartmouth (United States);
[c]Helmholtz Zentrum München GmbH (Germany)
[d]Dartmouth Hitchcock Medical Ctr. (United States)

ABSTRACT

A spectral analysis technique to enhance tumor contrast during breast conserving surgery is proposed. A set of 29 surgically-excised breast tissues have been imaged in local reflectance geometry. Measures of broadband reflectance are directly analyzed using Principle Component Analysis (PCA), on a per sample basis, to extract areas of maximal spectral variation. A dynamic selection threshold has been applied to obtain the final number of principal components, accounting for inter-patient variability. A blind separation technique based on Independent Component Analysis (ICA) is then applied to extract diagnostically powerful results. ICA application reveals that the behavior of one independent component highly correlates with the pathologic diagnosis and it surpasses the contrast obtained using empirical models. Moreover, blind detection characteristics (no training, no comparisons with training reference data) and no need for parameterization makes the automated diagnosis simple and time efficient, favoring its translation to the clinical practice. Correlation coefficient with model-based results up to 0.91 has been achieved.

Keywords: breast tumor; optical reflectance; optical scattering parameters; Principal Component Analysis (PCA); Independent Component Analysis (ICA).

1. INTRODUCTION

Cancer continues being one of the more frequent causes of death all over the world. Particularly breast cancer is the most common one among women [1]. In spite of its considerably high incidence, its survival rate is pretty good when it is detected in an early stage. That is the reason why WHO (World Health Organization) promotes the development of effective breast cancer prevention and control. Once a malignant tumor has been discovered, breast conserving therapy (BCT), which includes local surgery and radiation therapy, is the standard of care for early invasive breast cancers [2]. If the malignant tissue is only partially excised, the probability of recurrence is strong and the patient is recalled for a secondary surgery. On the other hand, a mastectomy can be too aggressive and problematic for a patient and even unnecessary when extracting a small and localized tumor. Actually BCT has been demonstrated to be equally safe and effective as mastectomy for most patients when surgical margins are clear of residual disease [3] [4]. Subsequently, a standardized approach is needed to discriminate between healthy and tumor tissues in the surgical margins [5].

In this study, statistical feature extraction techniques are used to explore the spatial distribution of tissue scattering and how its natural heterogeneity relates to tissue morphology. A scanning *in situ* spectroscopy platform has been designed to selectively sample the reflectance response of breast cancer specimens with the ultimate goal of developing a diagnostic adjunct to surgical procedures. However, the spectral response itself is too dense to interpret directly, and must be reduced into a few parameters that represent a diagnosis. In previous studies [5] [6], data reduction was achieved using an empirical approximation to Mie theory. In this study, Blind Source Separation (BSS) techniques will be used performing spatial analysis on spectral images. Techniques such as Principal Component Analysis (PCA) to reduce the dimension of data and Independent Component Analysis (ICA) to extract diagnostically useful features have been employed. The goal is to separate regions of interest in a fast, usable and accurate way. Results will be correlated with pathologist diagnosis.

Biomedical Applications of Light Scattering VI, edited by Adam P. Wax, Vadim Backman, Proc. of SPIE
Vol. 8230, 823010 · © 2012 SPIE · CCC code: 1605-7422/12/$18 · doi: 10.1117/12.909348

Several studies have already used ICA to extract image features on cancer detection and other biomedical applications [7]-[10]. Maps of probability of tumor have been taken from the ICA of RGB fluorescence images coming from the skin [7]. Kopriva et al. [7] concluded that ICA is a powerful method to be used also in other multi-channel medical imaging modalities. ICA is also used in magnetic resonance imaging. Chai et al. combined ICA and SVM methods to greatly improve the classification performance of brain tissues over other current techniques. Recently, Glatz et al. [8] have successfully applied PCA-ICA on biomedical multispectral imaging to separate the presence of different chromophores, using PCA to reduce dimension and ICA to finally divide the critical features on different images. ICA has also been applied on breast tissue characterization. Alrubaiee et al. [9] introduced the information theory based approach for the optical reconstruction of breast tissue specimens. They used multisource illumination to finally acquire the signals with multidetector which would be analyzed with ICA process. The reconstructed images were in good agreement with dimensions and pathological findings. Conceição et al. [10] used ICA on Microwave Imaging to extract important features for classification.

The main goal of this study is to improve the classification power of model-based parameters employed in previous works and to avoid preliminary training or knowledge about the samples for diagnostic classification. Finally, the discrimination performance of the blind detection will be compared with the tissue discrimination ability obtained from model-based parameters as the scattering power from tissue.

2. MATERIALS AND METHODS

2.1 Localized reflectance spectroscopy

A custom-built micro-sampling reflectance spectral system, developed in a previous study [6], images fresh breast tissue specimens. The localized reflectance imaging system consists of a confocal spectroscopic set-up and a raster-scanning sample platform built using translation stages. Each tissue sample was mounted on a glass slide and hydrated with a phosphate buffer solution. The system employs a quasi-confocal illumination and detection to constrain the overlapping illumination and detection spot sizes to 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 of the optical system was acquired, $R_{bkgrd}(\lambda)$, and subtracted from measured spectra, $R_{acquired}(\lambda)$, and the data was normalized, $R(\lambda)$, to the instrumental spectral response of the system using a Spectralon reference, (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 was used to describe the measured reflectance spectrum, R(λ), from a volume-averaged region of tissue [5]. Additionally, a Beer's Law attenuation factor is required to correct for the presence of significant local absorption by hemoglobin (Eq.2).

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

Parameters A and b are scattering amplitude and scattering power, respectively. Both depend on the size and number density of scattering centers in the volume of interrogated tissue, thereby reflecting variations in breast tissue morphology. Γ refers to the mean optical pathlength (dependent on the illumination and detection geometry), parameter $[HbT]$ is the total hemoglobin concentration, parameter 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, [11]). Oxygenated and deoxygenated hemoglobin were the dominant breast tissue chromophores encountered in the measured waveband [5]. Measured reflectance spectra were fit to this model using a nonlinear least squares solver to obtain estimates of scattering amplitude and scattering power relative to Spectralon. A measure of average scatter irradiance, I_{avg}, was calculated by integrating the reflectance spectrum over a waveband between 620 nm and 785 nm that avoids the hemoglobin absorption peaks.

In a previous study [12] the scattering power parameter, b, showed a good tissue discrimination capability being the best among all the scattering parameters. That is the reason why the present study will be focused in how a blind independent component analysis will correlate with the scattering power discrimination ability and with the pathologist diagnosis.

The present study is focused on the analysis of 29 specimens of breast tissue [5], with 48 different Regions of Interest (ROIs) corresponding to 7 different pathologies that are aggregated into 3 main categories: non-malignant, malignant and adipose. All the data is summarized in Table I.

Table I. Information on the statistics of the analyzed breast tissues.

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

2.2 Reduction of dimension of data with PCA

Principal Component Analysis (PCA) is usually employed to reduce the number of variables in set of data without losing any information, and also as a possible way to make these new variables more meaningful for the particular problem [13]. Several applications need PCA to determine the real dimension. In a dataset of m variables maybe that just $k < m$ variables would be needed to completely represent the same set. These variables can be initially correlated between them and after PCA they would be projected in a new space where the new projections would be uncorrelated. This is a second order method as its implementation only requires the information contained in the covariance matrix of the input data, which is a second order statistic. This fact makes the algorithm computationally simple and only dependent on matrix algebra equations.

PCA assumes linear mix of data, so a matrix of mixture $\mathbf{W}$, size $m \times m$, can be defined:

$$\mathbf{y} = \mathbf{W}\,\mathbf{x} \tag{3}$$

where $\mathbf{y}$ would be the $m \times n$ variable containing the uncorrelated components and $\mathbf{x}$ the $m \times n$ input data to the algorithm. On this study, m would be the number of wavelengths, 512, and n the number of pixels (observations).

To reduce the dimension, only $k < m$ uncorrelated projections would be kept and there must be a criterion to decide how many of them. The goal is to maintain as much variability of data as possible, so a measure can be the kept variance based on the sum of the eigenvalues of the selected uncorrelated components:

$$\frac{\sum_{q=1}^{k} D[\mathrm{q},\mathrm{q}]}{\sum_{q=1}^{m} D[\mathrm{q},\mathrm{q}]} 100(\%) \geq \text{threshold} \tag{4}$$

where D would be the diagonal eigenvalues matrix ($D[\mathrm{q},\mathrm{q}] = \lambda_q$).

On this particular case, m would be the 512 spectroscopy reflectance images that the system provides and the aim of PCA is to keep just $k < m$ uncorrelated images, therefore a threshold must be chosen. The kept variance presents different slopes for each different sample, which is illustrated on Figure 1. As a result, a different threshold has to be selected each time. This "dynamic threshold" was designed according to the increment of kept variance curve: when the variation from one kept component q from the next $q+1$ is less than 0.1% then the process of selection stops and the final uncorrelated components are those so far. From the initial 512 spectral images per tissue sample, no more than 5 components per sample were needed to be kept on this study. This means that less than 1% of the initial quantity of data would be needed, conserving almost all the information, which enormously reduces the need of memory and storage. The 99% rejected would be associated to redundancy and noise along each wavelength.

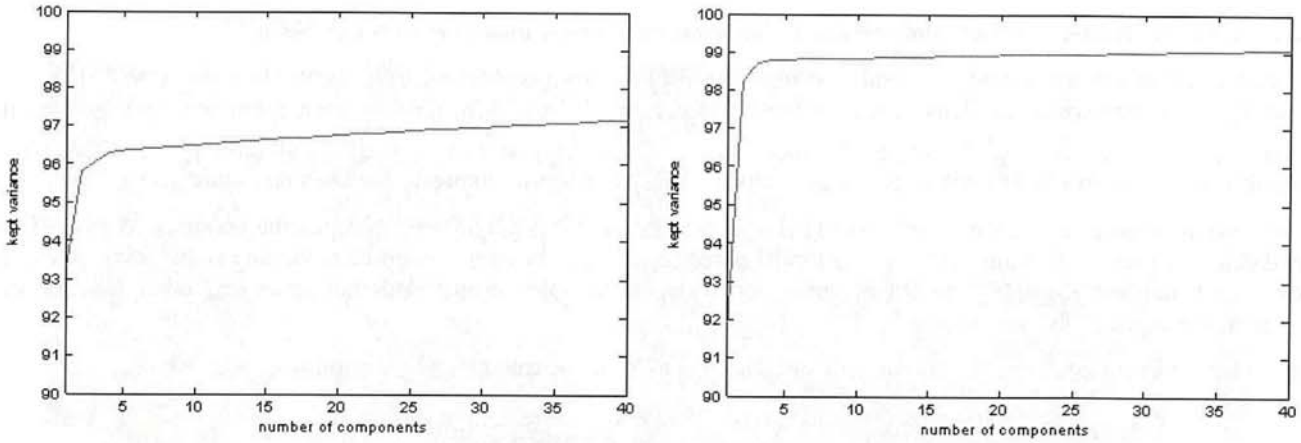


Figure 1. Evolution of the kept variance for two different tissue samples.

2.3 Extraction of diagnostic features with ICA

The Independent Component Analysis (ICA) is also employed to find a new subspace on a linear mix [13], so equation (3) could be also used to model being now m the number of principal components kept (n would continue being the number of pixel observations). However, in this case, output components $\mathbf{y}$ (usually called sources) would be independent, and not only uncorrelated. This idea is illustrated on Figure 2. Independence is, generally, a much restrictive characteristic, which would make possible a more diagnostically quality on resulting images. This means that some interested features of pathology classification could be still linearly mixed on the uncorrelated PCA results. This is the reason why ICA is applied to the output component from PCA. ICA requires high order statistics such as kurtosis, a four statistical moment that indicates the 'peakedness' of a distribution of probability.

Fast-ICA algorithm [14] has been selected for this study as it is computationally fast and does not need much memory space. Its purpose is to find independent components based on the Theorem of Central Limit establishment, which says linear mixes of independent variables tend to be Gaussian [15]. Thus, Fast-ICA extracts the more non-gaussian projections as possible from input data. Kurtosis is null for a Gaussian distribution, i.e., the way Fast-ICA extracts the sources is maximizing the kurtosis.

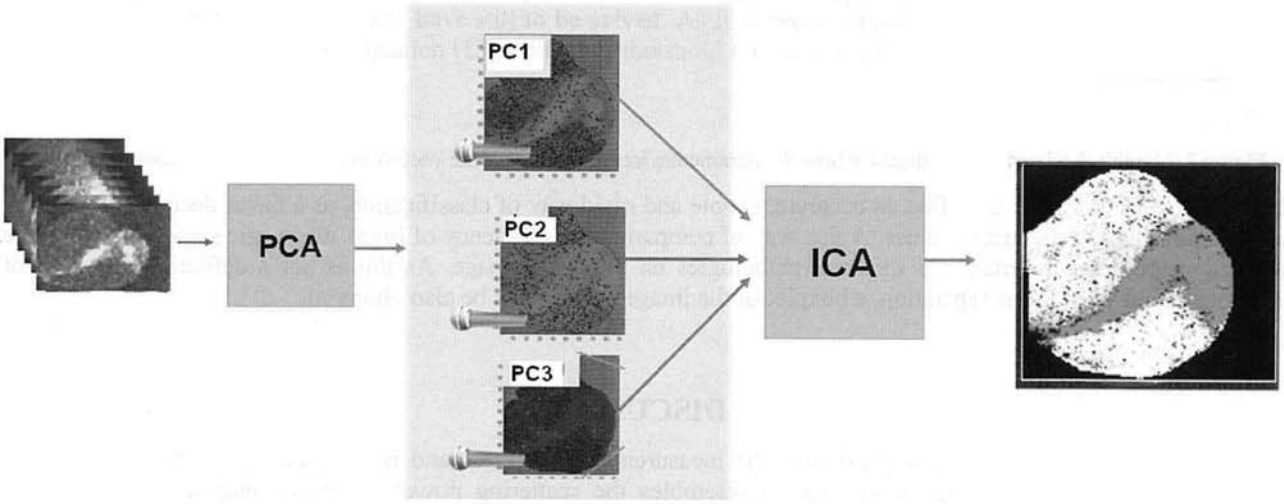


Figure 2. PCA-ICA application over the reflectance measured maps.

2.4 Linear correlation and distance between diagnostic categories: measures of comparison

With the purpose of comparing the results of this work with the images obtained from the model based ones [5] (A, b and I_{avg}), the Pearson's coefficient between these and the new PCA-ICA results has been computed. One dimension correlation has been checked, as images are processed in a 1D vector, placing row after row all those pixels related with tissue, i.e. no background or masked pixels are included. This process is graphically sketched on Figure 3.

The Pearson's linear correlation coefficient [16] is a measure of lineal relationship between the variables. If PCA-ICA had obtained exactly the same results as the model based process this Pearson's coefficient would result 1 or -1. A value close to 0 denotes absolutely no linear connection between variables, which does not reject any other function of relationship between them.

The Pearson coefficient, r , for two random variables X and Y can be calculated by the following expression:

$$r = \frac{Cov(X,Y)}{S_x S_y} = \frac{\sum_{t=1}^{n}(X-\overline{X})(Y-Y)}{\sqrt{\sum_{t=1}^{n}(X-\overline{X})^2}\sqrt{\sum_{t=1}^{n}(Y-\overline{Y})^2}} \quad (5)$$

where $Cov(X,Y)$ is the covariance between X and Y, $S_x S_y$ are their standard deviations, $\overline{X}$ and $\overline{Y}$ are the mean values and n is the number of observations (image pixels on this study).

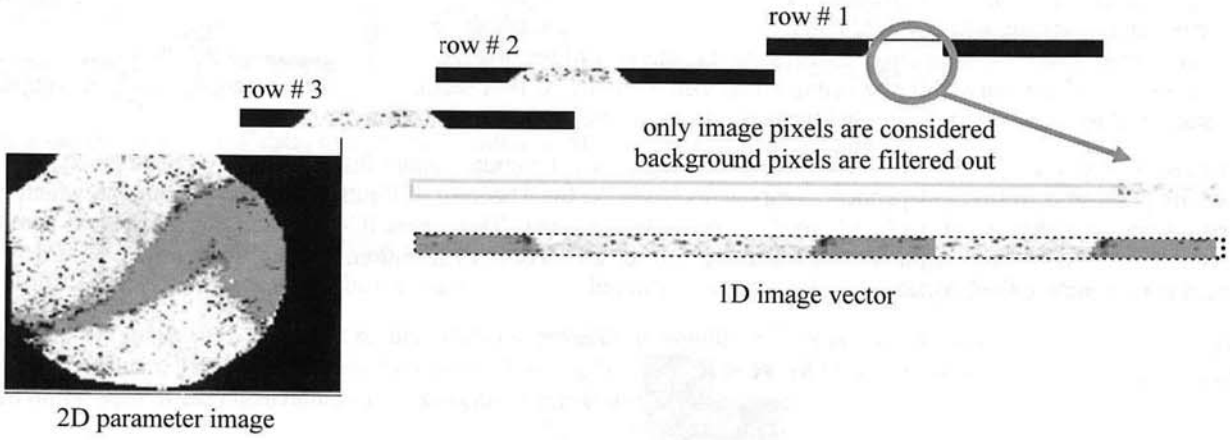
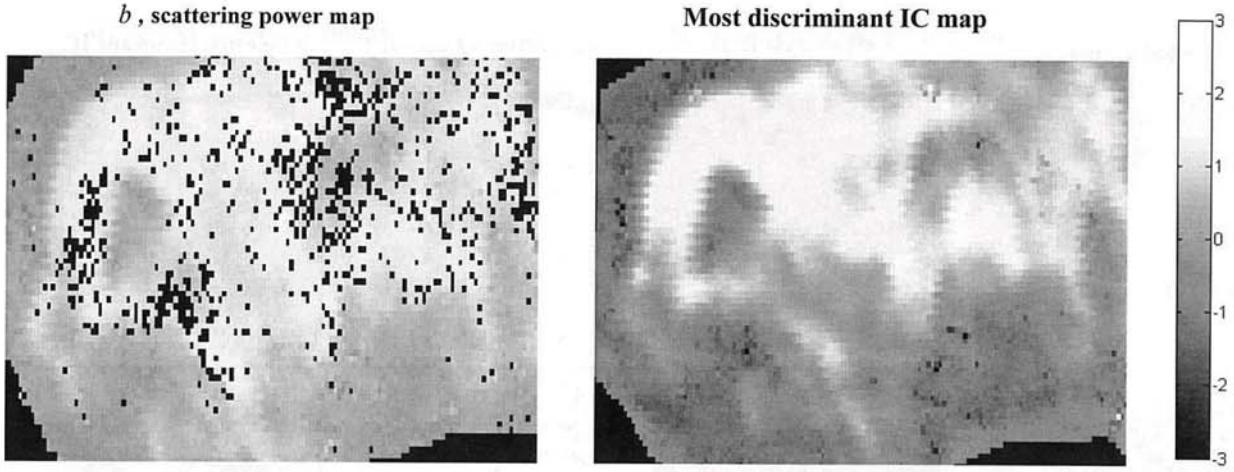


Figure 3. Graphical schematic that displays how the parameters image maps become vectorized for the 1D correlation analysis.

The final goal of this study is to find an accurate, simple and rapid way of classification, as a linear decision, depending just on a threshold of the image values. A fast way of comparing the efficiency of linear diagnostic separation would be the distance between the means of different pathologies on the same image. As this is not a definitive approach of checking the power of linear separation, a boxplot of the image points must be also observed.

3. DISCUSSION

The PCA-ICA process is directly applied over the measurements of broadband reflectance. Although no scattering parameter modeling is used, PCA-ICA solution resembles the scattering power 'b' image map on several tissue samples, but in a completely blind way. Actually, the mean Pearson coefficient value along the 29 samples dataset is 0.54. It could be preliminary concluded that ICA itself would be able to finally extract a scattering measure from the reflectance data. Furthermore PCA-ICA results seem to be consistent with pathologist H&E (Hematoxylin and Eosin stain) diagnosis.



Pearson coefficient, **r** (b, most discriminant IC) = **0.91**

Figure 4. The Pearson correlation between the scattering power map and the most discriminant IC map becomes really high in some specific tissue samples. Black points on 'b' (left image) correspond to masked points, those that did not fit on the scattering model [5].

However, PCA-ICA results do not always correlate so high with the scattering model-based map. Pearson's coefficients up to 0.91 have been achieved (as observed on Figure 4), thought in some cases this coefficient resulted nearly 0. Nevertheless, the image based only in the scattering power map was not always consistent with pathologist diagnosis, so the goal is not to achieve high correlation with previous model-based result but with the pathologist diagnosis map. Actually, we have observed cases were PCA-ICA image seemed to be diagnostically better than the model-based image even though Pearson coefficient between them is not very high. Figure 5 shows cases where the model-based result and the PCA-ICA result highly correlate between them and with pathologist areas (2nd column of Figure 5) but also cases where they not correlate between them but, even so PCA-ICA is consistent with pathological regions.

PCA-ICA also seemed to be more powerful on a linear separation than previous model. While the distance between mean values of different pathologies on model-based images is 0.9, the distance on PCA-ICA images reaches 1.97. Although this information may not conclude that PCA-ICA is preferred, the boxplots displayed in Figure 6 shown that the resulting PCA-ICA diagnosis were more separable. Even though the boxplots from PCA-ICA seem to behave in a more favorable way, some problems have still to be solved. As ICA process presents some inherent ambiguities with sign and order due to the model equation [15] a different threshold for each sample had to be chosen.

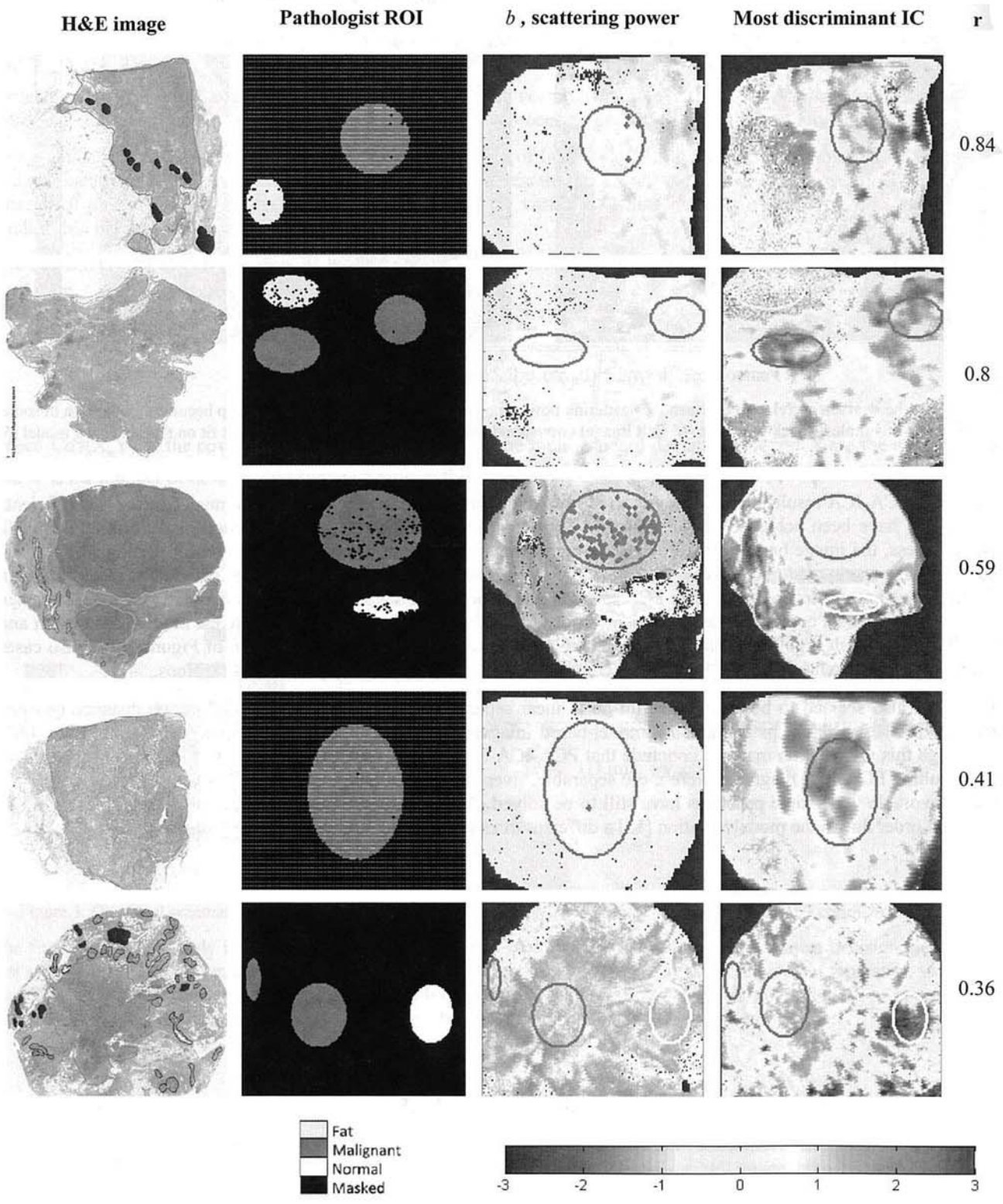


Figure 5. Visual correlation between H&E slides (1st column), pathologist ROI's (2nd column), scattering power map - *b* (3rd column) and the most discriminant independent component image (4th column). The 5th column shows the correlation coefficient between the scattering power map and the most discriminant component image map.

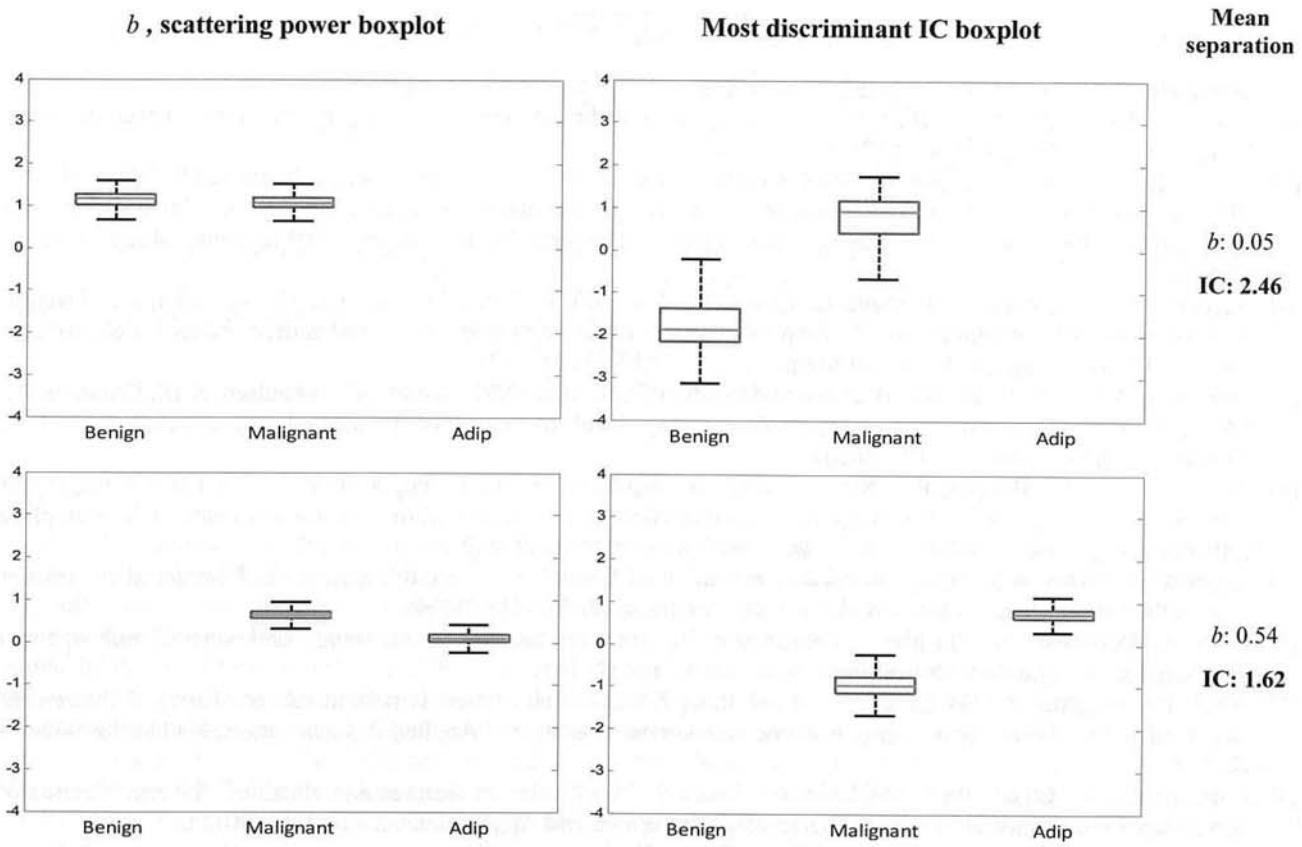


Figure 6. Separation between diagnostic categories is improved when the IC component is considered.

4. CONCLUSIONS

ICA blind separation technique has been applied after extracting, with PCA, uncorrelated components of optical broadband reflectance. No models, no trainings, no comparation with distances have been used, which makes the technique computationally fast and simple. Fast-ICA algorithm has been used for the source extraction process. This fast implementation makes the whole system very suitable for providing a rapid, pathologic diagnosis during surgical procedures.

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. In addition, the results mostly correlate with the model-based image obtained from the scattering power from tissue. This would mean somehow the parameters from model could be obtained in a blind way.

In spite of the fact that this extraction of independent features suggest a linear way of classification, which is much more efficient that other classifiers, the final threshold should have to be different for each sample, as inherent ambiguities of ICA analysis still have to be solved. This will be the main goal of future works on this line of research as once this ambiguities are solved a linear classifier will be able to finally be proved.

ACKNOWLEDGEMENTS

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

REFERENCES

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

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

[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] 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).

[5] 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).

[6] 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).

[7] Kopriva, I., Peršin, A., "Unsupervised decomposition of low-intensity low-dimensional multi-spectral fluorescent images for tumour demarcation," Medical Image Analysis, 13, 507-518 (2009).

[8] Glatz, J., Deliolanis, N., Buehler, A., Razansky, D., Ntziachristos, V., "Blind source unmixing in multi-spectral optoacoustic tomography," Optics Express, 19, 3175-3184 (2011).

[9] Alrubaiee, M., Xu, M., Gayen, S.K., Alfano, R.R., "Localization and cross section reconstruction of fluorescent targets in ex vivo breast tissue using independent component analysis," Applied Physics Letters, 89(13), 133902-1-3 (2006).

[10] Conceição, R.C., O'Halloran, M., Glavin, M., Jones, E., "Evaluation of features and classifiers for classification of early-stage breast cancer," Journal of Electromagnetic Waves and Applications, 25(1), 1-14 (2011).

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

[12] 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).

[13] Semmlow, J.L., [Biosignal and Biomedical Image Processing. MATLAB-Based Applications], CRC Press (2004).

[14] Hyvärinen, A., Oja, E., "A fast fixed-point algorithm for independent component analysis," Neural Computation 9(7), 1483-1492 (1997).

[15] Hyvärinen, A., "Survey on Independent Component Analysis," Neural Computing Surveys 2, 94-128 (1999).

[16] Setchi, R. et al., [Knowledge-based and intelligent information and engineering systems], Lecture Notes in Computer Science, Springer (2010).