Blind breast tissue diagnosis using independent component analysis of localized backscattering response

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 ZentrumMünchen GmbH (Germany)
[d]Dartmouth Hitchcock Medical Ctr. (United States);

ABSTRACT

A blind separation technique based on Independent Component Analysis (ICA) is proposed for breast tumor delineation and pathologic diagnosis. Tissue morphology is determined by fitting local measures of tissue reflectance to a Mie theory approximation, parameterizing the scattering power, scattering amplitude and average scattering irradiance. ICA is applied on the scattering parameters by spatial analysis using the Fast ICA method to extract more determinant features for an accurate diagnostic. Neither training, nor comparisons with reference parameters are required. Tissue diagnosis is provided directly following ICA application to the scattering parameter images. Surgically resected breast tissues were imaged and identified by a pathologist. Three different tissue pathologies were identified in 29 samples and classified as not-malignant, malignant and adipose. Scatter plot analysis of both ICA results and optical parameters where obtained. ICA subtle ameliorates those cases where optical parameter's scatter plots were not linearly separable. Furthermore, observing the mixing matrix of the ICA, it can be decided when the optical parameters themselves are diagnostically powerful. Moreover, contrast maps provided by ICA correlate with the pathologic diagnosis. The time response of the diagnostic strategy is therefore enhanced comparing with complex classifiers, enabling near real-time assessment of pathology during breast-conserving surgery.

Keywords: breast tumor; localized backscattering; optical scattering parameters; Independent Component Analysis (ICA).

1. INTRODUCTION

Early detection of breast cancer is the keystone of recovery [1]. Once a localized malignant tumor is detected, a surgery procedure must follow. Breast-Conserving Therapy (BCT) has been the standard of care for early invasive breast cancers and it includes local excision and radiation treatment to the breast [2]. However, if tumor is not extracted correctly and some malignant cells are left on the patient, there is a high risk of reappearance, even with the radiation treatment [3]. Drastic solution for surgeons could be a mastectomy, i.e., complete excision of the breast, but this intervention produces serious consequences on patients, not only biologically but also psychologically [4] since it a dramatic fact in a social and emotional way. All this problems may not appear if the extraction of malignant tissue is accurate. To this aim, a high resolution and fast system of microscopy should be installed on the surgery room to correctly delineate boundaries of malignancy with precision. This would be the ultimate goal of the study presented on this line of research.

A scanning spectroscopy platform is proposed to analyze the spatial distribution of the scattering because it reflects the natural heterogeneity observed in tissue morphology [5]. To analyze only the first order scattering, a modeling of spectral data has to be done to minimize multiple scattering and absorption effects. Scattering parameters have previously been used on breast cancer detection. Wang et al. [6] developed a method of image reconstruction of Mie scattering parameters with near-infrared source, interpreting adipose and fibrograndular tissues on different breast tissue samples. Also non-optical scattering has been studied. Recently, Elshemey et al. [7] evaluated the diagnostic capability of X-ray scattering profile parameters on breast cancer excised samples, achieving a diagnostic accuracy up to 97%.

Biomedical Applications of Light Scattering VI, edited by Adam P. Wax, Vadim Backman, Proc. of SPIE
Vol. 8230, 823014 · © 2012 SPIE · CCC code: 1605-7422/12/$18 · doi: 10.1117/12.909326

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/13/2013 Terms of Use: http://spiedl.org/terms

On a previous work [5] a k-NN classifier demonstrated discrimination between benign and malignant tissues with a sensitivity and specificity of 91 and 77% respectively, using the scattering parameters as starting points. Nevertheless this classifier is hard to train and computationally slow compared with much simpler classifiers as those based on the evaluation of linear thresholds. In this study, Independent Component Analysis (ICA) has been applied over the scattering parameters to extract more diagnostically defining features from the model-based ones.

ICA has been previously used on breast cancer detection and other biomedical application involving spectral analysis. ICA is applied on breast cancer detection to obtain interesting features on mammograms [8]-[10]. Amara et al. experimental results [8] showed 71.2% of accuracy using ICA to detect suspicious regions on mammograms. Costa et al. [9] combined ICA and Linear Discriminant Analysis (LDA) reaching a 95.2% of accuracy discriminating benign and malignant masses on mammograms. Kopriva et al. [10] extracted tumor maps from skin fluorescence RGB images with ICA and other unsupervised linear mixture model. They concluded that these techniques would also work in other multi-channel medical imaging systems.

This study analyses the viability of independent component analysis when applied on scattering parameters with the aim of simplifying and improving the classification of malignancy. The resulting extracted ICA features seem to be easily classified with a threshold providing a fast and simple linear classification.

2. MATERIALS AND METHODS

1.1 Localized reflectance measurements

The imaging system consists of a confocal spectroscopic set-up and a raster-scanning platform built using translation stages. Both subsystems, optical and electromechanical, are integrated within a custom developed LabVIEW interface. Figure 1 shows a schematic of the measurement set-up. The illumination optics consists of a 50 μm core fiber (F1) coupled to a 100W tungsten-halogen white light source placed at the front focal point of an achromatic lens (L1). A 10X, long working distance, plan-apochromatic objective (L2) was used to refocus the light onto the sample. The illumination optics was modeled in Zemax software to assure an illumination spot size smaller than one scattering length (typically 100 μm) over the entire wavelength band (510-785 nm) that covers the strong hemoglobin absorption peaks. The detection optics uses the same objective to collect the backscattered light from the sample and a 50/50 beam splitter to separate illumination and detection beam paths. Another achromatic lens (L3) focuses the detected photons onto a 100 μm core optical fiber acting as the confocal pinhole. A more detailed description of the system and its calibration and characterization procedures can be found in [11].

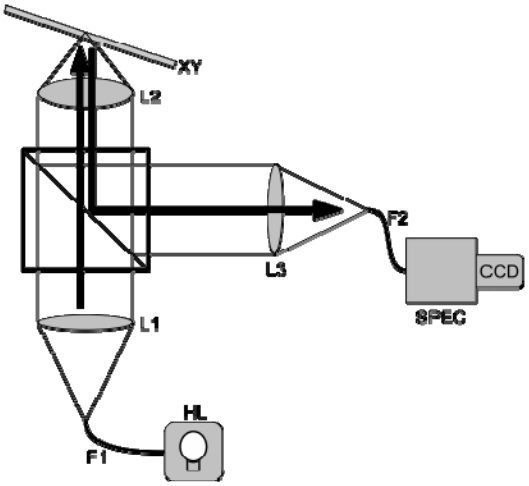

Fig. 1. Scatter imaging system set-up.

The acquired spectral reflectance is spectrally corrected versus a Spectralon reference (Labsphere, Inc., North Sutton, New Hampshire) to allow direct comparison between tissue samples. Afterwards, the spectrally corrected reflectance is

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/13/2013 Terms of Use: http://spiedl.org/terms

fitted by an empirical approximation to Mie's theory. This approximation accounts for scatter versus wavelength and a Beer's Law attenuation factor is included to correct for the presence of significant local absorption by hemoglobin:

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

Parameters A and b are scattering amplitude and scattering power, respectively. Both magnitudes 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, [12]). Oxygenated and deoxygenated hemoglobin were the dominant tissue chromophores encountered in the measured waveband [5]. The measured reflectance spectra are fit to this model using a nonlinear least squares solver to obtain estimates of scattering amplitude and scattering power relative to the Spectralon reference. Apart from the specific scattering parameters, amplitude (A) and power (b), a measure of average scattering irradiance, I_{avg}, is calculated by integrating the reflectance spectrum over a waveband between 620 nm and 785 nm that avoids the hemoglobin absorption peaks.

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 typologies: 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**

1.2 Independent Component Analysis

The hypothesis of this work is that hidden and diagnostically powerful factors could be linearly mixed within the scattering parameters maps: A, b and I_{avg}. This hidden features could be more classifiable in a more rapid and simple way. The model faced on this analysis is shown on Eq. (2).

$$\mathbf{y} = \mathbf{W}\,\mathbf{x} \quad (2)$$

where $\mathbf{y}$ is a $m \times n$ variable containing the output independent components; $\mathbf{x}$ is the $m \times n$ input variable, i.e. the scattering parameter maps (A, b and I_{avg}); $\mathbf{W}$ is the $m \times m$ mixture matrix. On this study m would be three (the considered scattering optical parameters) and n the number of pixels analyzed on each image. The unknown latent variables contained in $\mathbf{y}$ are supposed to be mutually independent and non-Gaussian [13].

Fast-ICA method [14] was selected because of its simplicity and computational behavior. Fast-ICA algorithm is based on the maximization of the kurtosis, related to the 'peakeness' of the probability distribution function, of the output components in order to reach the maximum non-Gaussianity. The reason for this is based on the Theorem of Central Limit establishment, which says that linear mixtures of independent variables tend to be Gaussian, so the non-gaussianity of components of mixture will denote independency. The formulated problem on Eq. (2) has some inherent ambiguities with sign and order of the output components, or sources, which may cause difficulties on selecting a global threshold on a final linear classification.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/13/2013 Terms of Use: http://spiedl.org/terms

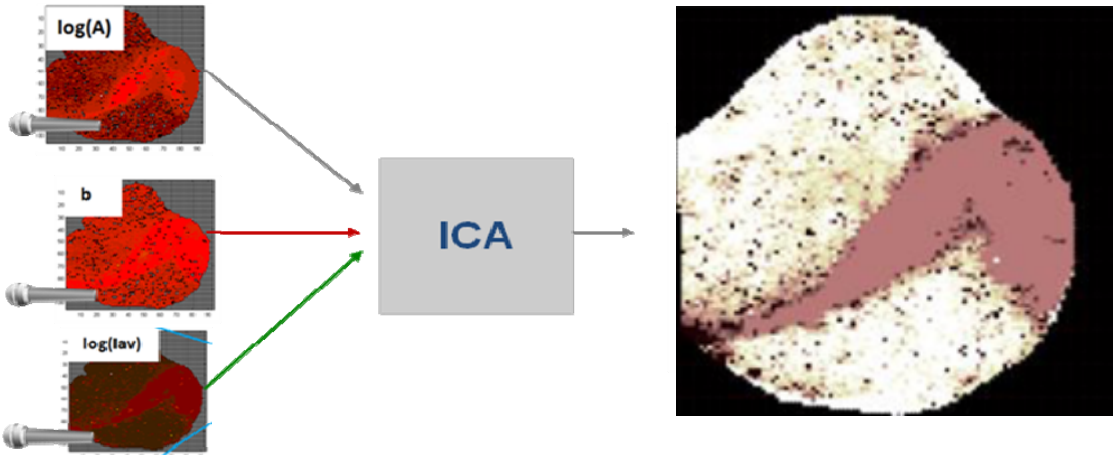

Fig. 2. ICA "cocktail party" [13] application over the scattering parameters: amplitude (A) , power (b) and average scattering irradiance, I_{avg} .

1.3 Scatter plot

Strong correlation was observed on previous study [5] between the scattering power (b) and the logarithm of the scattering amplitude (A). This statistically similar behavior of both parameters made ICA to just reveal two independent components as output. Then we achieve two sets of variables. A useful technique to represent two variables is a 2D scatter or dispersion plot, where one variable behaves as abscise and the other one as ordinate. The resulting plane shows the relationship between variables, i.e. how separable they are and which way of classification would be the more accurate.

3. DISCUSSION

Given the strong statistical similarity between scattering power and amplitude, and assuming that this was the reason why FastICA did not generate three different output components, the work has been focused only on the scattering power, b, and on the average scattering irradiance, I_{avg} . As result, the analysis generates two independent components, IC_1 and IC_2 , from these two model-based parameters b and I_{avg} . At each image pixel, output (IC_1 and IC_2) and input data (b and I_{avg}) will be now linked by the mixture matrix $\mathbf{W}$ as stated in Eq. (3).

$$\begin{bmatrix} \boldsymbol{IC}_1 \\ \boldsymbol{IC}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{w}_{11} & \boldsymbol{w}_{12} \\ \boldsymbol{w}_{21} & \boldsymbol{w}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{I}_{avg} \end{bmatrix} \tag{3}$$

After a detailed analysis of the scatter plots of the ICA results, some samples exhibit a strong correlation with the behavior of the scatter plots of optical parameters. Figure 3 shows how the ICA only rotates the axes, fact that becomes useless from a classification point of view. In these cases, the corresponding mixing matrix $\mathbf{W}$ defined on Eq. (2) and (3) results to be almost diagonal or anti-diagonal concluding that IC image maps are very similar to the optical parameters map and no improved relevant information is provided after ICA application. A simple diagnosis criteria based on the evaluation of a threshold in just one of the scatter plot axes is able to be implemented in both cases, scattering parameters domain and IC components domain.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/13/2013 Terms of Use: http://spiedl.org/terms

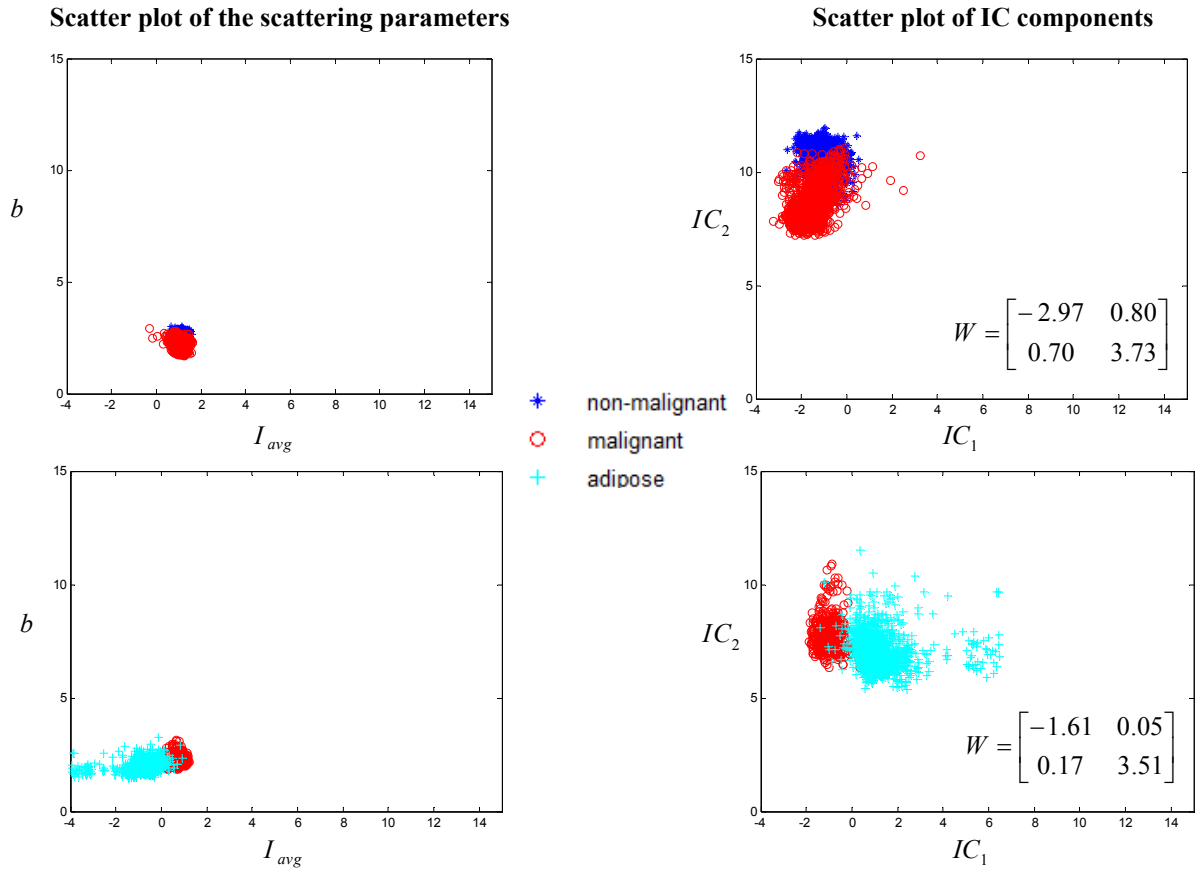

Fig. 3. Scatter plots of scattering parameters (left) and independent components (right) on samples where the mixture matrix is almost diagonal or anti-diagonal.

Nevertheless, on some other tissue samples, a subtle improvement of diagnosis separation is achieved on the scatter plots provided by ICA. This fact would make possible a better linear classification through the evaluation of a linear threshold. The right column of Figure 4 shows how the separation of clusters, corresponding to different tissue diagnosis, is improved after ICA implementation. This fact can be observed on the rotation and expansion of the data that would help in the definition of the diagnosis threshold.

This fact validates the idea that the improvement is not only on the implementation of a simple lineal classifier for diagnosis, but also on the possibility of using just one independent component to visually guide the surgeon. As the variance of the data relies only on one axis direction, all the diagnostically interested information on the optical parameters could be collected on a single image corresponding to one of the independent components. Figure 4 suggests that the evaluation of just one independent component, the vertical one (IC_2), would provide guidance to the surgeon for tissue discrimination. The insets of Figure 4 show that the corresponding mixing matrixes $\mathbf{W}$ are not now diagonal for these cases. This analysis reveals that the final IC image becomes a mixture of both optimal parameters (b and I_{avg}) extracting all the relevant information from both of them.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/13/2013 Terms of Use: http://spiedl.org/terms

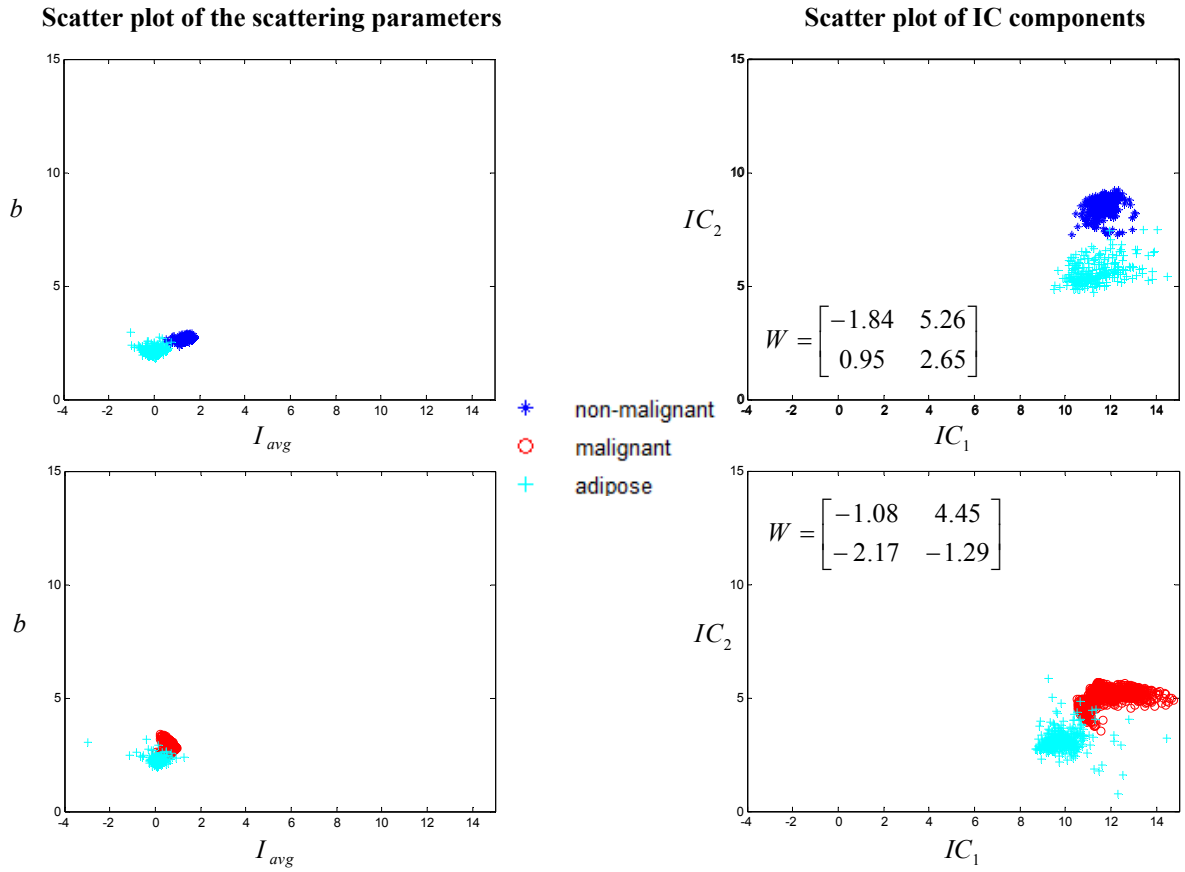

Fig. 4. Scatter plots of scattering parameters (left) and independent components (right) on tissue samples where the mixture matrix is not diagonal or anti-diagonal.

When the most discriminant independent component is displayed (right column of Figure 5), high visual correlation between pathological regions of interest and different contrast levels on the image was achieved. Furthermore, this independent component map is consistent with the H&E image (Hematoxylin and Eosin stain) supplied by the pathologist (left column of Figure 5).

Finally, the most significant independent component provides a more accurate and contrasted visual result than the optical parameters themselves. The metric demonstrating this fact is the mixing matrix **W**, as already described. As on most of the samples this matrix is not diagonal, an algebraic mix of optical parameters in performed on ICA result, as shown on Eq. (4):

$$IC_i = w_{i1}\, b + w_{i2}\, I_{avg} \tag{4}$$

ICA would be getting the most discriminate features from each one of the scattering parameters generating an improvement. In addition, this mixture map provides more contrast than the parameters' maps themselves, as Fig. 6 displays, and could provide a clearer guidance map for the surgeon when he is in an intraoperative context.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/13/2013 Terms of Use: http://spiedl.org/terms

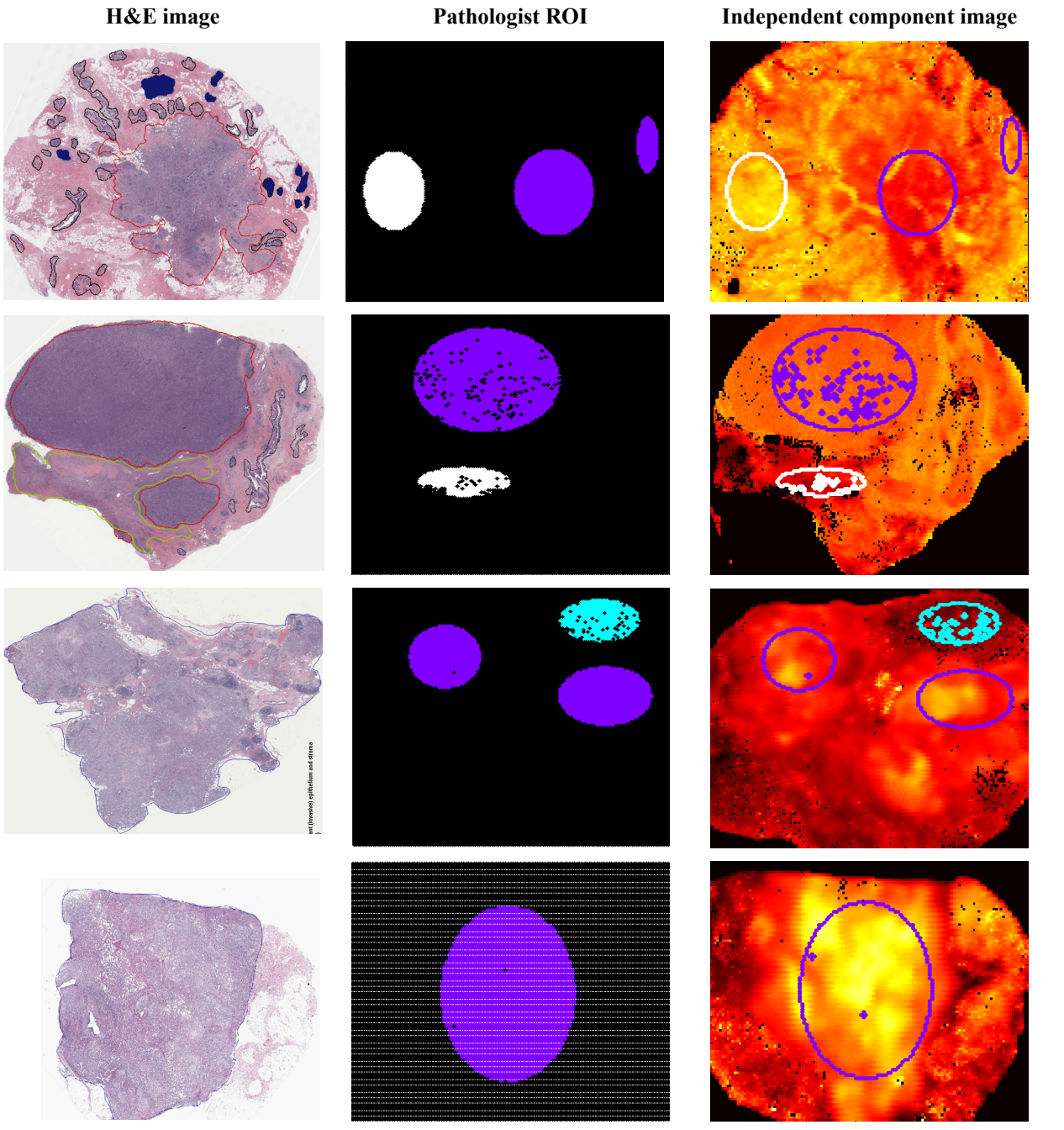

Fig. 5. Visual correlation between H&E slides (left), pathological regions of interest (center) and the different levels on most discriminant independent component image (right). The same color is not always associated to same pathology due to ICA inherent ambiguities.

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/13/2013 Terms of Use: http://spiedl.org/terms

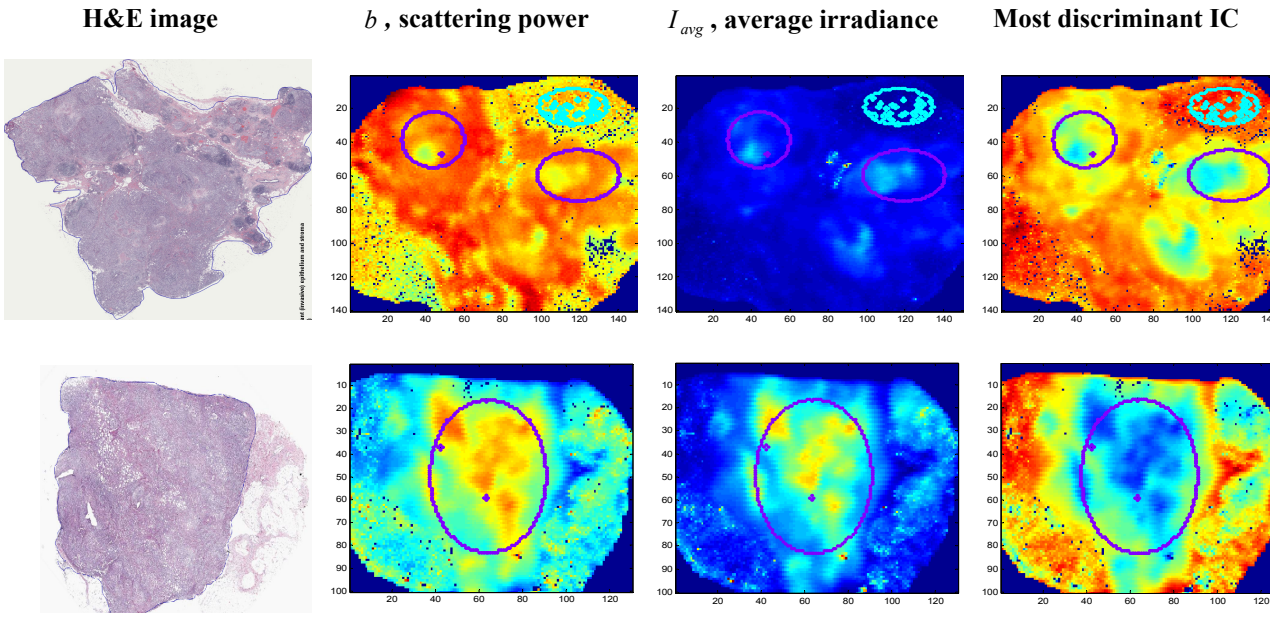

Fig. 6. Visual correlation between H&E slides (1st column), scattering power map (2nd column), average scattering irradiance (3rd column) and the most discriminant independent component image (4th column).

4. CONCLUSIONS

Spatial domain ICA technique has been applied over optical scattering parameter model-based maps, extracted from localized reflectance measures of breast tissue resected specimens. Although three scattering parameters (*A*, *b* and *Iavg*) where firstly used as ICA input, finally scattering amplitude *'A'* was avoided due to its statistical similarity with the scattering power *'b'*. As advantage, to help in the following automatic classification stage, the two resulting ICs have been represented in a scatter/dispersion plot. As a result of this post-processing stage, it has been checked that ICA subtly helps in the grouping of the diagnostic categories in tissue samples where the scattering parameter *b* and *Iavg* were less separable. This fact can be checked on the mixture matrix of the definition of the problem, **W**. After displaying the ICA resulting tissue image, consistency with pathology regions of interest and visual correlation with H&E images has been achieved. Furthermore, ICA seems to favorably mix the diagnostic information from *b* and *Iavg* to finally obtain a more discriminate surgeon-guide image.

However some problems with inherent ambiguities of ICA algorithm still have to be solved. Also, it has been checked that sometimes ICA does not change the optical parameters, as the mix matrix **W** is mostly diagonal or anti-diagonal. This fact may be suggesting that optical parameters are already independently enough on some tissue samples, which motivates the idea of applying ICA directly on the reflectance measures to avoid modeling, and some significant maps could be obtained in a absolutely blindly way.

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).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/13/2013 Terms of Use: http://spiedl.org/terms

[3] Smitt, M.C., Nowels, K.W., Zdeblick, M.J., Jeffrey, S., Carlson, R.W., Stockdale, F.E., Goffinet, D.R., "The Importance of the Lumpectomy Surgical Margin Status in Long-Term Results of Breast-Conservation," Cancer 76(2), 259-267 (1995).
[4] Sociedad Española de Oncología Médica: http://www.seom.org.
[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] Wang, X., Pogue, B.W., Jiang, S., Dehghani, H., Song, X., Srinivasan, S., Brooksby, B.A., Paulsen, K.D., Kogel, C., Poplack, S.P., Wells, W.A., "Image reconstruction of effective Mie scattering parameters of breast tissue in vivo with near-infrared tomography," Journal of Biomedical Optics 11(4), 041106 (2006).
[7] Elshemey, W.M., Desouky, O.S., Fekry, M.M., Talaat, S.M., Elsayed, A.A., "The diagnostic capability of x-ray scattering parameters for the characterization of breast cancer," Medical Physics 37(8), 4257-4265 (2010).
[8] Abu-Mara, F., Abdel-Qader, I., "Detection of breast cancer using independent component analysis," 2007 IEEE International Conference on Electro/Information Technology, EIT 2007, Conference Digest 428-431 (2007).
[9] Costa, D.D., Campos, L.F., Barros, A.K., Silva, A.C., "Independent component analysis in breast tissues mammograms images classification using LDA and SVM," Proceedings of the IEEE/EMBS Region 8 International Conference on Information Technology Applications in Biomedicine, 231-234 (2008).
[10] 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).
[11] 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).
[12] Jacques, S.L., Prahl, S., Oregon Medical Laser Center. 2010.
[13] Hyvärinen, A., "Survey on Independent Component Analysis," Neural Computing Surveys 2, 94-128 (1999).
[14] Hyvärinen, A., Oja, E., "A fast fixed-point algorithm for independent component analysis," Neural Computation 9(7), 1483-1492 (1997).

Downloaded From: http://proceedings.spiedigitallibrary.org/ on 06/13/2013 Terms of Use: http://spiedl.org/terms