ICA-GUIDED DELINEATION OF BREAST CANCER PATHOLOGY

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

A surgeon-guided independent component analysis from optical reflectance measurements is proposed for breast tumor delineation. Independent Component Analysis is first applied to extract the most relevant features from local measures of broadband reflectance and then a tumor probability indicator is obtained and provided utilizing surgeon assistance to resolve the inherent ambiguities in the independent component calculation. A set of 29 breast tissue samples have been diagnosed achieving a sensitivity of 90.57%, and specificity of 93.98%.

Index Terms— breast tumor delineation, Independent Component Analysis (ICA), Principal Component Analysis (PCA), optical scattering.

1. INTRODUCTION

Breast cancer incidence is much higher than any other cancer among women. Fortunately it has also a very favorable relation incidence-mortality as early detection remains the cornerstone of breast cancer control [1].

Following the early detection, a careful selection of the treatment must be done. 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 it has been demonstrated to be equally safe and effective as mastectomy for most patients when surgical margins are clear of residual disease [2]. Nevertheless, surgical margin evaluation lacks of standarization, which risks patient's long-term prognosis.

A scanning spectroscopy system has been designed with the purpose of indentifying residual cancer in surgical excision specimens. The scattering spectrum provides information about the size and number density of cellular components and structural matrix elements [3]. The optical system provides 512 different scattering maps (one per spectral band). Principal Component Analysis (PCA) is initially used to reduce the spectral dimension of the optical data. Then, Independent Component Analysis (ICA) is applied to blindly extract the most important features for diagnosis [4].

ICA analysis has been previously validated in breast cancer detection applications [5-6]. ICA has obtained classification features on mammograms [5]. PCA employment as a preprocessing step reduces the problem dimension and provides better performance when compared with the application of individual algorithms [6]. Maps of tumor probability have been extracted from the ICA of RGB fluorescence images taken from the skin [7]. Kopriva et al. believe the method is prone to be used in other multi-channel medical imaging modalities. Recently, Glatz et al. [8] have successfully applied PCA-ICA on multispectral optoacoustic tomography to blindly unmix distinct exogenous fluorophores.

This paper employs ICA as a diagnosis predictor, previously optimized by PCA, to aid in the discrimination of the different tissue pathologies present at the surface of breast surgical specimens, which could be finally implemented as an optical biopsy in an intraoperative context. As ICA presents ambiguity problems in recovering the sign of the sources of the sample, a procedure based on the selection of "well-known" points based on surgeon selection is proposed. Results have been compared with regions of interest (ROI) evaluated by the pathologist to final validation of the classification.

2. OPTICAL DATA

The scatter imaging system consisted of a weakly confocal spectroscopic system with illumination and detection spot sizes smaller than one mean scattering length (typically 100 μm for tissue [9], i.e., one pixel covers 100 μm x100 μm sample surface), and a raster-scanning platform built using linear translation stages. The instrument operates in the 510 nm – 785 nm spectral waveband with a broadband fiber-coupled tungsten-halogen light source. An optical-fiber 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. A schematic and a more detailed description of the system can be found in this previous paper [10].

We have analyzed 29 tissue samples, with 48 different Regions of Interest (ROIs) 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. MULTIVARIATE ANALYSIS

The optical system provides 512 scattering images. PCA application reduces this data set into uncorrelated images. But some interesting factors that could be more representative for the diagnosis could still be hidden or latent if source uncorrelation is solely sought. ICA finds sources that satisfy a stronger condition as statistical independence and, consequently, it can aid in highlighting this hidden relevant information.

3.1. PCA

The information contained in the 512-wavelength dataset is quite redundant and this redundancy can be removed employing PCA to reduce the spectral dimension, i.e. the number of images. Each tissue sample $\mathbf{x}$ is a $M \times N$ matrix, where M is the number of wavelengths (512 data dimension) and N the number of pixels on the tissue sample (about 4000 pixels, observations). The aim is to find a new data projection after a linear transformation where the new projections, $\boldsymbol{s}$, are uncorrelated.

$$\mathbf{s}_{M\times N} = \mathbf{W}_{M\times M}\,\mathbf{x}_{M\times N} \tag{1}$$

The covariance matrix from x, $\mathbf{C}$, must be calculated, after mean-centering of the data. The eigenvector matrix is the one that diagonalizes the covariance matrix, i.e.,

$$\mathbf{V}^{-1}\mathbf{C}\mathbf{V} = \mathbf{D} \tag{2}$$

where $\mathbf{D}$ a diagonal matrix containing the eigenvalues of $\mathbf{C}$. The mixing matrix, $\mathbf{W}$, is then given by:

$$\mathbf{W}_{M\times M} = \mathbf{V}^{*}_{M\times M} \tag{3}$$

A criterion is established to reduce the number of maintained uncorrelated components from M to L. The chosen criterion is to maintain $L<M$ eigenvalues with a joint variance above a specific threshold, as shown in (4).

$$\frac{\sum_{q=1}^{L} D(q,q)}{\sum_{q=1}^{M} D(q,q)} 100(\%) \geq \text{threshold} \tag{4}$$

As the kept variance presents different slopes depending on the tissue sample, a *dynamic threshold* based on the derivative of the kept variance curve has been selected. To this end, the maintained components will be those whose kept variance varies more than 0.1% from the one obtained in the previous step. The resulting number of maintained uncorrelated components uses to vary from 2 to 5, being usually 4. Figure 1 show two different cumulative variance plots; the first few components (left, in shaded area) correspond to the very large eigenvalues, while the many components to the right have small eigenvalues and are presumed to be noise.

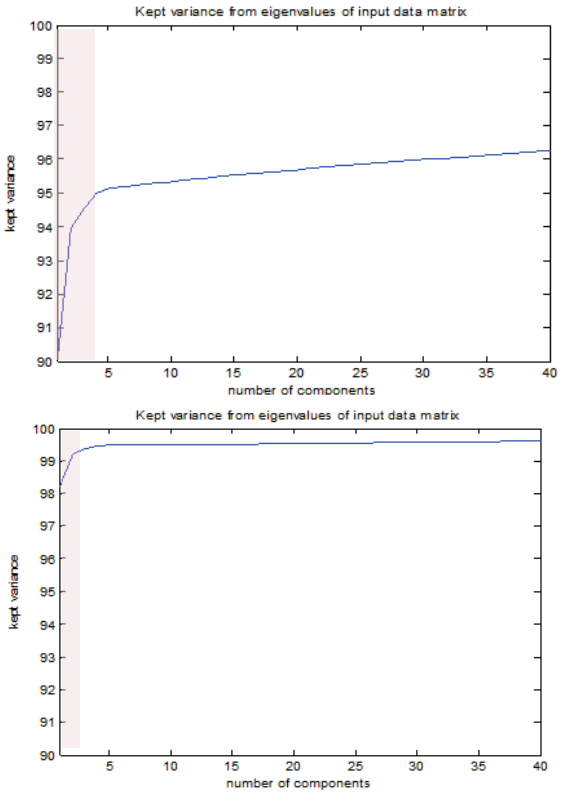


Fig. 1. Different samples require a different number of kept components through a dynamic threshold evaluation.

3.2. ICA

Some interesting factors for diagnosis might be still linearly mixed on the uncorrelated PCA maintained components.

A similar situation as in equation (1) is faced, but now **x** are the uncorrelated components chosen from PCA analysis, and ***s*** components, sources, are supposed to be diagnostically determinant. In this study, it is very important to emphasize that the sign of the components of matrix **W** could be inverted without affecting the model. This generates ambiguity in the sign of the diagnostics results.

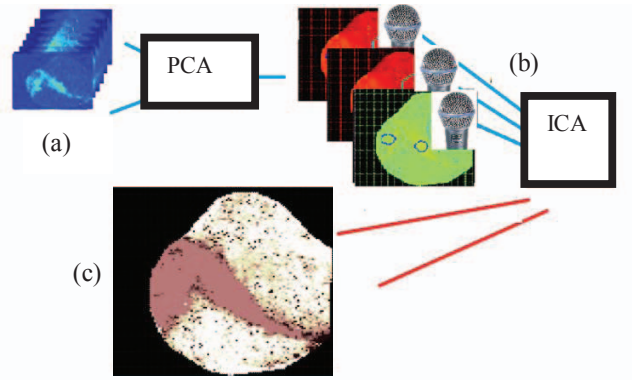


Fig. 2. PCA-ICA implementation for the analysis of the scattering from breast tissue samples. (a) 512 input images; (b) uncorrelated images, reduced dimension; (c) final diagnosis map

Fast-ICA algorithm [4] has been used for the ICA analysis. It is based on maximization or minimization of the fourth statistical moment, i.e. kurtosis. Kurtosis is null for a Gaussian distribution and provides a measure of the "peakedness" of the probability distribution of the data. The purpose is to look for the more non-Gaussian components possible. The Theorem of Central Limits establishes that the linear mix of independent variables tends to be gaussian. Thus, the non-gaussianity is the key of estimating the mixing matrix **W** of these random variables. Furthermore, Fast-ICA is computationally simple and requires little memory space. Our studies could prove that a successful diagnosis map is provided by the independent component space.

Moreover, this procedure is totally blind since no training, no modeling, no supervision or no previous signal information is needed. Figure 2 represents how the PCA-ICA analysis of the breast tissue scattering problem has been implemented in this work and its correlation with the typical "cocktail party" problem.

The diagnostic maps obtained after the PCA-ICA implementation show different signs to different pathologies. When malignant and benign/adipose tissues are presented on the sample, each category has a different sign. If only adipose and benign tissues are present, then they also have different signs. Tissue pathologies become separated on the image.

4. WELL-KNOWN POINTS STRATEGY

Although tissue regions become separated by ICA, the ambiguities of ICA algorithm generate several difficulties to the absolutely blind selection.

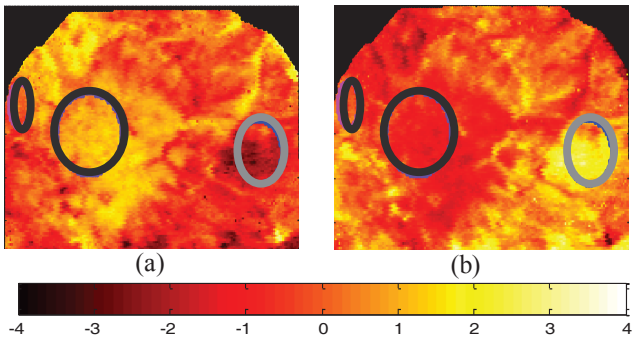


Fig. 3. Two different results for the same tissue sample. On (a) the IC of malignant tissue (black circles) is positive and benign tissue (gray circle) is negative, but they exchange signs on (b).

The same IC sign is not always associated to the same pathology on different repetitions over the same sample, or on the same tissue pathology for different patients. Figure 3 represents the same tissue sample with the superimposed ROIs diagnosed by the pathologist. Ambiguities of ICA with sign create difficulties to establish a malignancy threshold.

Nevertheless this ambiguity does not have to be a problem on the context of this application. The goal is not to identify the malignancy, but its extent on the tissue surface. This extra and supervised information can be added to the process as extra "well-known" points the surgeon can easily identify. This allows ICA to identify which tissue is being separated. As a simulation of this surgeon selection, 25 pixels on each ROI have been selected as recognized, contrasted points to be certainly diagnosed. Then a detection mask can be easily created.

5. RESULTS AND DISCUSSION

Once the surgeon identifies the tissues in the image, the generation of a diagnosis mask is provided by ICA because it separates different pathologies, assigning them different signs. After validation with pathologist ROIs, sensitivity and specificity can be computed on a pixel basis. The mean sensitivity of malignancy detection has been 90.57%, while the specificity has reached 93.98%. Figure 4 shows the delineation results on two different samples.

A more conservative interpretation of results could be given by a map showing the tumor probability, Figure 5. A scale transformation of the IC's values into the (0,1) range provides an alternative vision of margin delineation: 1 for the presence of tumor and 0 for a null probability of tumor.

(a) White light image (b) H&E section (c) Pathologist ROIs (d) PCA-ICA result

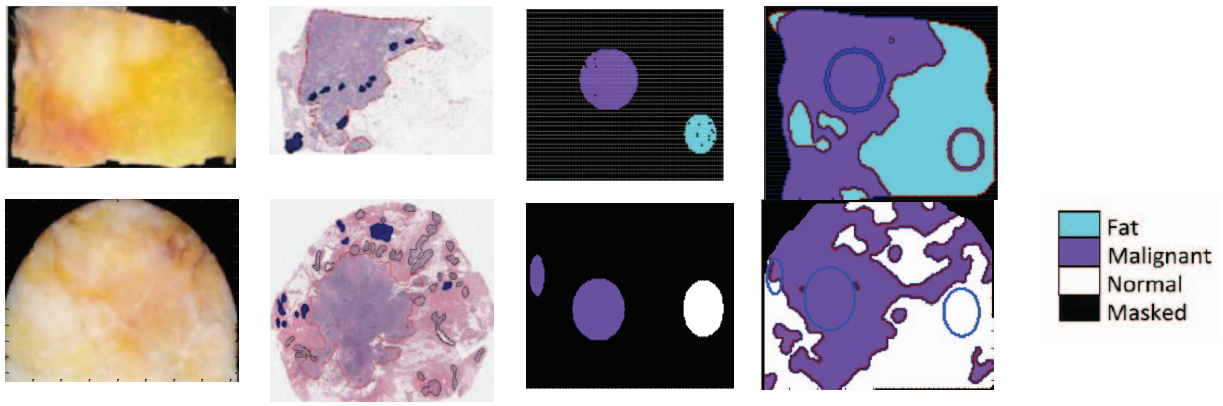


Fig. 4. PCA-ICA results spatially smoothed (d) compared with their associated white light image (a), H&E section (b) and pathologist diagnosis (c).

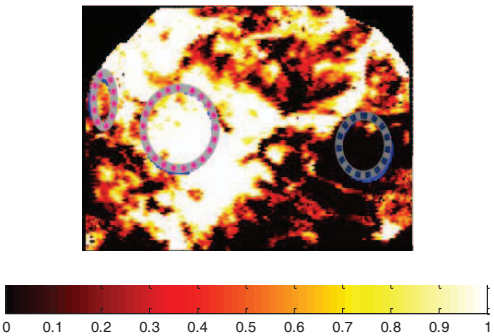


Fig. 5. Probability of tumor map: 1 indicates presence of tumor, 0 indicates benign tissue. Pathologist ROIs malignant (magenta point circles) and benign (blue squares circle) are also shown.

6. CONCLUSIONS

PCA and ICA have been subsequently applied over optical reflectance images. PCA reduces the dataset dimension maintaining the relevant information, whereas ICA provides a significant diagnosis image map. Ambiguities with the sign of the ICs prevent the determination of a fixed threshold for a linear classification, but a well-known set of pixels could be provided by the surgeon to solve it

Sensitivity and specificity up to 90% have been achieved within the ROIs used for the evaluation. However, significant correlation with H&E maps is also obtained, which suggest that a future application of the system could be margin delimitation. The goal of this approach is not to diagnose malignancy, but to evaluate the extent of tumor between two different tissue types in the image field. Cooperation of surgeon is needed during the BCT surgery. Surgeon's supervision is followed by a simple classification, and a computationally efficient ICA algorithm to obtain the diagnosis image.

ACKNOLEDGEMENTS

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] 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] S.Srinivasan, B.W.Pogue, S.Jiang, H.Dehghani, C.Kogel, S.Soho, J.J.Gibson, T.D.Tosteson, S.P.Poplack, K.D.Paulsen, "Interpreting hemoglobin and water concentration, oxygen saturation, and scattering measured in vivo by near-infrared breast tomography," *PNAS*, 100(21), 12349-12354, 2003

[4] A. Hyvärinen and E. Oja. "A Fast Fixed-Point Algorithm for Independent Component Analysis". *Neural Computation*, 9(7):1483-1492, 1997.

[5] R.Gallardo-Caballero, C.J.García-Orellana, M.Macías-Macías, H.M.González-Velasco, F.J.Lypez-Aligué, "Independent component analysis applied to breast cancer detection on digitized mammograms," *InW. Congr. Ser.*, 1281(5), 1052-1057, 2005.

[6] F.Abu-Amara, I.Abdel-Qader, "Detection of breast cancer using independent component analysis," *IEEE International Conference on Electro/Information Technology,* 428-431, 2007.

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

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

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

[10] A.M. Laughney, V. Krishnaswamy, P.B. Garcia Allende, O.M.Conde, W.A. Wells, K.D. Paulsen, B.W. Pogue, "Automated classification of breast pathology using local measures of broadband reflectance", Journal of Biomedical Optics, Vol. 15, No. 6, pp. 066019-066019-12, 2010.