UNIVERSITY
OF TRENTO

DIPARTIMENTO DI INGEGNERIA E SCIENZA DELL'INFORMAZIONE

38123 Povo – Trento (Italy), Via Sommarive 14
http://www.disi.unitn.it

AN INTEGRATION BETWEEN SVM CLASSIFIERS AND MULTI-RESOLUTION TECHNIQUES FOR EARLY BREAST CANCER DETECTION

P. Rocca, F. Viani, M. Donelli, M. Benedetti, and A. Massa

January 2011

Technical Report # DISI-11-202

An integration between SVM Classifiers and Multi-Resolution Techniques for Early Breast Cancer Detection

P. Rocca*, F. Viani, M. Donelli, M. Benedetti, and A. Massa
ELEDIA Group – Department of Information Engineering and Computer Science
University of Trento, Via Sommarive 14, I-38050 Trento, Italy
E-mail: andrea.massa@ing.unitn.it, Web-page: http://eledia.ing.unitn.it

Introduction

Because of the high contrast between the dielectric properties of normal and malignant breast tissues at microwave frequencies, microwave imaging techniques seem to be very attractive diagnosis methods for cancer detection [1][2]. In such a framework, inverse scattering methods are very promising tools, but their practical application is strongly limited by the need of 3D reconstructions, high spatial resolutions, and fast processing. Recently, to reduce the high computational costs and to fit the real-time requirements, inversion methods based on learning by example techniques have been proposed [3]. LBE approaches based on support vector machines (SVMs) [3] and neural networks (NNs) [4] have been satisfactorily applied in various and complex electromagnetic problems. When dealing with breast cancer detection, the inversion process is recast as a classification or regression problem where the unknowns are retrieved from the data (i.e., the electric field samples collected in an external observation domain) by approximating the unknown relation data-unknowns through an off-line data fitting procedure (training phase). Once the training procedure (performed once and off-line) is completed, the characteristics of the malignant breast tissue are real-time estimated in the testing phase. In such a work, the detection problem is addressed by integrating a SVM-based classifier with an iterative multi-zooming procedure. More in detail, a succession of approximations of a probability map of the presence of pathology is determined. At each step, the spatial resolution of the risk-map is improved in a limited set of regions of interest (ROIs) defined at the previous zooming step and characterized by a greater value of the occurrence probability of a malignant tissue. The multi-step procedure is stopped when a stationary condition on the probability and on the number of ROIs is reached. The achievable trade-off between computational complexity and spatial resolution is preliminary assessed by discussing a selected set of numerical simulations concerned with both noiseless as well as corrupted data.

Mathematical Formulation

Let us consider the three-dimensional scenario shown in Fig. 1. The simplified human breast model, composed by a cylinder and a hemisphere of total height

H_D, constitutes the 3D investigation domain I_B. Both cylinder and hemisphere have radius R_D.

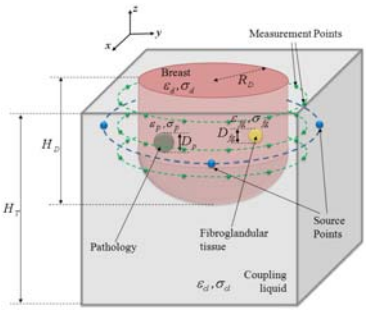

Figure 1 - Schematic of the three dimensional synthetic breast model

By considering a multi-view imaging system, L short dipoles are circularly arranged around the breast. They are immersed in a coupling medium with constitutive parameters similar to those of the normal breast tissue to maximize the coupled microwave energy. The scattered electromagnetic field $E_{scatt}^{(l,s,z_k)}(x,y,z)$ is collected on $K=3$ different circular layers with a set of S_k sensors per layer. Starting from the knowledge of the scattered signal, the problem of finding a pathology in a given region of the domain is recast as the definition of a probability risk-map. Towards this end, the 3D investigation domain is partitioned into C cubic cells. The state of the c-th cell can be either empty ($\chi_c=-1$) or occupied ($\chi_c=1$) if any or a pathology resides in the cell. The output of the classification problem consists of a cell occupation probability q_c, $c=1,...,C$. Such a problem is then solved by means of a SVM-based technique whose main steps are summarized in the following.

- **Step 0**: Training Phase. The classifier is off-line trained by assuming the knowledge of a set of N known input/output relations $\left\{\left[\underline{E}_{scatt}^{(l,s,z_k)},(x_m,\ y_m,\ z_m)\right]^n;\ n=1,...,N\right\}$, that relates the measured scattered field $E_{scatt}^{(l,s,z_k)}$ to the location (x_m,y_m,z_m) of the malignant tissue. According to the SVM theory, the classification problem is solved by determining the linear decision function $\hat{\Gamma}=\underline{w}\cdot\underline{\varphi}(E_{scatt}^{(l,s,z_k)}(x,y,z))+b$ defined in a higher dimensional feature space where the examples are mapped through the non-linear operator $\underline{\varphi}$. As a consequence, the problem unknowns are $\underline{w}$ and b. They are determined during the training phase;
- **Step 1:** Low-Order Estimation ($s=1$). The 3D investigation domain I_B is uniformly partitioned into C cubic sub-domains. Then a coarse risk-map, characterized by an homogeneous resolution level r, is computed by solving through SVM the arising classification problem;
- **Step 2:** Multistep Process ($s>1$). Starting from the risk-map reconstructed at the previous step, a set of reduced investigation domains ($RoIs$), where the pathologies are supposed to be located, is identified. Towards this end, the

previous map is binarized by thresholding and processed with standard image processing operators. The spatial resolution level $(r = r+1)$ is enhanced only in the $RoIs$ and a multi resolution risk-map is obtained by applying the SVM classification to the $RoIs$;

- **Step 3:** Termination Criterion ($s = S_{opt}$). If the dimensions and the number of $RoIs$ between two consecutive steps are unchanged, then the multi-resolution process is terminated.

Numerical Results

In order to point out the capabilities and current limitations of the proposed strategy, let us consider the following representative example chosen from a set of numerical test cases. The pathology has been located at $x_m = 0.0\lambda, y_m = 0.0\lambda, z_m = 0.116\lambda$, with a diameter of $D_p = 0.016\lambda$ and constitutive parameters $\varepsilon_p = 50$, $\sigma_p = 1.6\ S/m$.

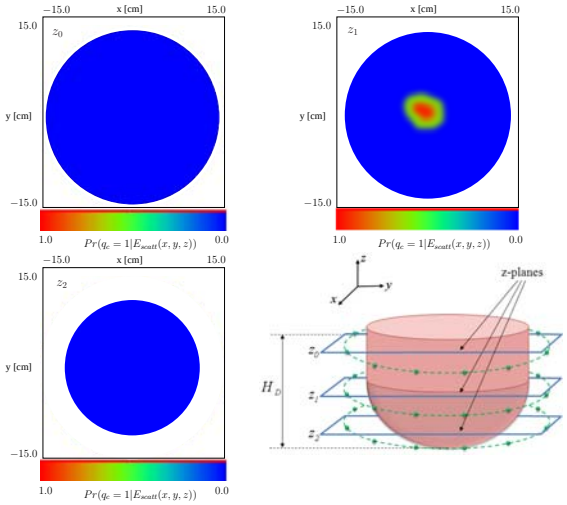

Figure 2 - Risk-map slices at three different z-planes

To simulate a more realistic breast specimen, a fibroglandular tissue inclusion of $D_{fg} = 0.014\lambda$ in diameter and dielectric properties $\varepsilon_{fg} = 32.7$ and $\sigma_{fg} = 1.28\ S/m$ has been randomly positioned in I_B. The investigation domain of height $H_D = 0.75\lambda$ and radius $R_D = 0.5\lambda$, is characterized by dielectric parameters $\varepsilon_d = 10$ and $\sigma_d = 0.4\ S/m$. The breast model has been supposed to be immersed in a cubic tank of side length $H_T = 2\lambda$ and filled with a coupling liquid whose characteristics are $\varepsilon_{cl} = \varepsilon_d = 10$ and $\sigma_{cl} = \sigma_d = 0.4\ S/m$. Moreover, the scenario has been probed by $L = 4$ short dipoles operating at $f = 1.0\,GHz$. The positions of the electromagnetic sources were $\theta_l = (l-1)(2\pi/L)$, where $\theta_1 = 0°$, while the scattered electric field has been collected at $S_k = 15; k = 1,...,K; K = 3$ measurement points disposed on three different circumferences $R_s = 0.55\lambda$ in radius and located at

the planes $z_k = 0.35\lambda - (0.25\lambda \times k);\ k = 0,...,K-1$. The training set has been generated by considering a set of $N = 200$ input/output relations and different positions and dimensions of the pathology. For completeness, the numerical data have been corrupted by adding a Gaussian noise ($SNR = 30\,dB$). Concerning the multistep procedure, the $RoIs$ have been partitioned at each step in $C_{(r)} = 10 \times 10 \times 10,\ r = 1,...,R$ cuboids. Figure 2 shows three slices of the 3D risk-map obtained at the end of the iterative process ($s = S_{opt} = 2$). As it can be observed, the pathology is correctly localized although the region with higher probability values slightly overestimates the actual dimension of the tumor. Moreover, it should be noticed that only the presence of the pathology has been estimated although a fibroglandular artifact exists.

Concerning the computational burden, the determination of the multi-resolution risk-map obtained at the end of the IMSA-SVM strategy requires about 800 msec.

Conclusions

In this work, a multi-scaling classification approach for breast cancer detection has been presented. The approach exploits the generalization capabilities of a SVM-based methodology and a multi-resolution strategy. In order to assess the effectiveness and the feasibility of the proposed technique, a preliminary result has been reported.

Acknowledgment

This research was carried out within the framework of a project entitled "Development of Advanced Automatic Analysis Methodologies for Environmental, Industrial and Biomedical Monitoring", funded by the Italian Ministry of Foreign Affairs.

References:

[1] M. El-Shenawee and E. L. Miller, "Spherical harmonics microwave algorithm for shape and location reconstruction of breast cancer tumor," *IEEE Trans. Med. Imag.*, vol. 25, no. 10, pp. 1258-1271, Oct. 2006.

[2] Y. Xie, B. Guo, L. Xu, J. Li, and P. Stoica, "Multistatic adaptive microwave imaging for early breast cancer detection," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 8, pp. 1647-1657, Aug. 2006.

[3] A. Massa, A. Boni, and M. Donelli, "A classification approach based on SVM for electromagnetic subsurface sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2084-2093, Sep. 2005.

[4] I. T. Rekanos, "Neural-network-based inverse-scattering technique for online microwave medical imaging," *IEEE Trans. Magn.*, vol. 38, no. 2, pp. 1061-1064, Mar. 2001.