Filtering by Approximated Densities applied to Texture Modelling for Mammography

Virginie Ruiz* & Anthony G. Constantinides**

*ISMRA-ENSI, GREYC CNRS URA 1526, 6 Bd Maréchal Juin, 14150 Caen Cedex, France
Tel: +33 31 45 27 05; fax: +33 31 45 26 98; e-mail: v.ruiz@greyc.ismra.fr
**Imperial College of Science, Technology and Medicine, Dept Electrical, Electronic Engineering Exhibition Road, London SW7 2BT, UK

ABSTRACT

Many techniques are currently used for breast abnormality location and breast cancer detection, in particular. One find statistical approaches involving second and/or higher order statistics. The applicability of the filtering by approximated densities (FAD) is here demonstrated. The FAD introduced to alleviate limitations due conventional Kalman modelling, is applied to texture modelling for mammography. This application uses the simplest form of FAD involving second order statistics.

1. INTRODUCTION

Mammographic image analysis constitutes a wide field of investigation for the signal processing application. Many techniques are currently used for breast abnormality location and breast cancer detection, in particular. In mammography, it is difficult to discern the appropriate features for purpose of tumour detection. Mammograms are among the most difficult images to interpret because contrasts of regions are often very low. As it is known, the sooner a tumour is diagnosed the greater are the chances to be cured, therefore, features indicative of a breast abnormality are often very small.

As far as the texture modelling is concerned, many techniques are applied. Among those, one find statistical approaches involving second and/or higher order statistics which leads us to the subject of this paper.

The filtering by approximated densities (FAD), introduced to alleviate limitations due to conventional Kalman modelling, is here presented in its first practical application to texture modelling in mammograms. The FAD is a non-linear, adaptive filtering technique. It uses a maximum entropy principal under linear constraints. The distributions created by applying the entropy principle are of an exponential type and they depend on a finite number of parameters. The FAD-method is essentially based on the development of a logarithm for the computation of prior and posterior probability density functions as linear combinations of several functions chosen according to some specific criterion. Our FAD concept is presented in section 2.

In order to implement the FAD for the breast tumour detection, we must extract a finite number of pertinent characteristics from grey-level histograms of mammograms, so as, we can elaborate functions of an exponential type for the definition of probability density functions that characterise what we call on the on hand 'healthy tissue', and on the other hand 'tumorous tissue'. This part is developed in section 3.

For this first application, the FAD is intentionally implemented (section 4) in its simplest form using Gaussian approximations.

In section 5., the segmentation we achieve clearly indicates the area of the tumour.

2. METHODOLOGY

In order to develop our approach based upon the concept of filtering by approximated densities, we start with the state equation

$$X(k+1)=f(\,X(k),\,W(k)\,) \qquad (1)$$

where f is any non-linear function and $W(k)$ is white state noise having any distribution.

This equation is used to compute through a transformation of variables, the linear constraints:

$$l_i = E[\varphi_i(X(k+1))]$$

upon the variable $X(k+1)$ when the measure $Y(1,...,k)$ is known. At time instant $k+1$, the information to be retained to describe the posterior state distribution is defined through the given functions φ_i, which can be indicator of sets in order to give the probability of the state to be in a given set or power functions in order to yield moments of the state variable, etc.

The state equation precisely defines the transformation of random variable. Hence, the expectation of a function φ is expressed as a double integral with respect to $X(k)$ and $W(k)$

$$l_i=E[\varphi_i(f(X(k),W(k)))]$$

The predicted distribution is the distribution which maximises the entropy under the linear constraints, l_i. Probability density function created, from the constrained maximisation of the entropy, is of an exponential type. The following general expression can be defined for the density of an exponential type statistical distribution:

$$f(x)=exp(\sum_i \lambda_i \varphi_i(x)) \quad \lambda_i \in R$$

It is observed from equation (1) that the dynamic behaviour of $X(k+1)$-state is not retained in its entirety. Indeed, due to the entropy principle, the method inherently approximates all probability distributions through exponential type density functions. A study of the state equation without noise reveals the properties of the asymptotic distribution. Its mean is located at the zeros of the equation $x=f(x,0)$. The probability density has several local maxima, i.e., it is multimodal.

We pay particular attention to the number of local maxima in the probability density in order to develop the logarithm of the density function from base functions φ_i, which must be chosen accordingly. The coefficients in these linear expressions vary with the family of densities under consideration and are computed at each update and prediction.

At each time instant k, the output equation is used

$$Y(k)=X(k)+V(k) \qquad (2)$$

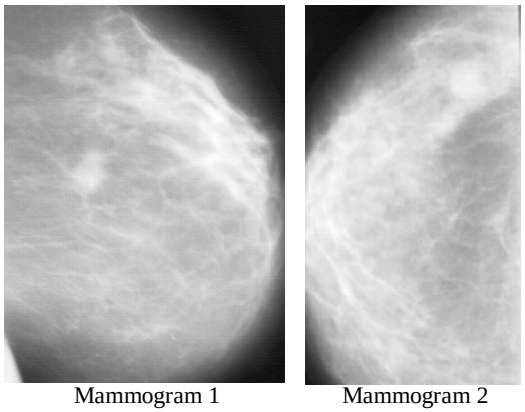
where $V(k)$ is white output noise with an exponential type density function.

The Bayes' formula can be applied to (2) when $X(k)$-state and $V(k)$-noise are of exponential type. Indeed, that enables the computation of the $X(k)$-state distribution conditional upon the observed output $Y(1,...,k)$, which remains as an exponential type as the output equation is linear.

3. TEXTURE CHARACTERISATION

Before implementing FAD for breast tumour location/detection, we must define a prior statistical distribution characterising on the one hand, 'healthy tissue' and one the other hand, 'tumorous tissue'. Hence, we must extract a finite number of pertinent characteristics from the mammograms 1 & 2:

Figure 1: Mammograms



Mammogram 1 Mammogram 2

When aiming for an early diagnosis, one knows tumours are very small and might not be detectable by human eyes. Within a mammogram a tumour commonly overlaps an area of 16x16-pixel (more or less). Hence, we define the region of support as a 16x16-pixel window. Thus, histograms should be relevant of healthy or tumorous areas and statistical characteristics extracted should be reliable.

Histograms upon areas presenting 'healthy tissue' provides grey-level distributions which are clearly non-Gaussian as opposed to tumorous areas that present grey-level distributions nearly Gaussian.

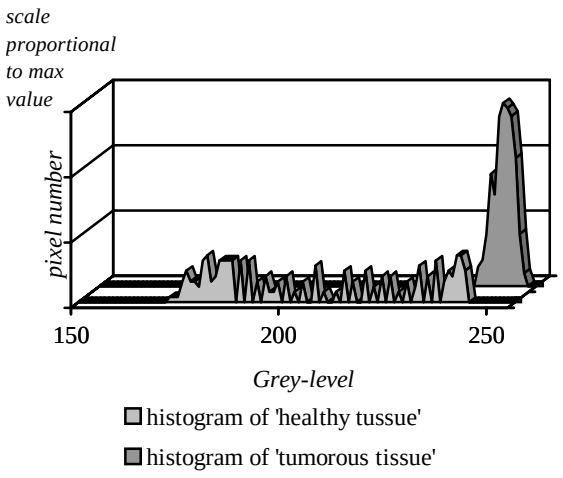
Figure 2., displays two histograms that can be said representative of healthy and tumorous cases.

Grey-level moments up to third order are significant and discriminating:

- 'Healthy tissue', the means are in the range (177,230), and standard-deviations are at least 5.

- 'Tumorous tissue', the means are beyond 230 and standard-deviations are less then 5.

Figure 2: Two typical histograms



As first approximation from values previously given, Gaussian distributions can be considered to characterised healthy and tumorous regions. We already can observed that 2nd-order approximation may prove itself insufficient to discriminate the two cases, when processing over mammary gland areas where the skewness value is the mere value that definitely separate the two cases under consideration.

Despite this later remark, the FAD is intentionally developed in its simplest form using Gaussian approximations. Hence, FAD only uses statistical characteristics up to second order.

According to our concept, we now define the prior density functions. A Gaussian distribution is of exponential type. The logarithm of its density function is linearly developed from the base functions: $\varphi_0(I)=1$, $\varphi_1(I)=I$, $\varphi_2(I)=I^2$, where I is pixel's grey-level.

The linear constraints l_i, $(i=0,1,2)$ are the expected values of the functions φ_i, $(i=0,1,2)$. Hence, these constraints correspond to moments up to second order of the Gaussian distribution. Thus, we set the prior distribution under each hypothesis:

- **Hypothesis H_1**, 'healthy tissue': Pixel's grey-level $I_{m,n}$ is taken to be Gaussian distributed with mean $m_1=210$, and standard-deviation $\sigma_1=10$. The density is then expresses as:

$$ {}^1f(I)=exp({}^1\lambda_0 + {}^1\lambda_1.I + {}^1\lambda_2.I^2) \qquad (3) $$

- **Hypothesis H_2**, 'tumorous tissue': The characteristics are the mean $m_2=240$ and standard-deviation $\sigma_2.=1$. The density function is defined through expression:

$$ {}^2f(I)=exp({}^2\lambda_0 + {}^2\lambda_1.I + {}^2\lambda_2.I^2) \qquad (4) $$

where Lagrange multipliers ${}^j\lambda_i$, $(j=1,2;\ i=0,1,2)$ are bijectively obtained from the linear constraints, i.e., the moments.

4. NON-LINEAR FILTERING

The filtering by approximated densities updates and predicts the posterior density functions conditional upon the observed values within a 16x16-pixel region of support, according to the state-space model define as follows with k denoting the k-th window:

$$ I_{k+1} = (1-a_j).I_k + b_j + W_k \qquad (5) $$
$$ Y_k = I_k + V_k \qquad (6) $$

where state noise W_k and output noise V_k are Gaussian distributed.

The state equation (5) introduces two constant coefficients a_j and b_j, $(j=1,2)$. Their numerical values are such that, under each hypothesis, the equation (5) yields a limit state I_∞ with statistical characteristics whose values are the prior ones.

The detection/update use the output equation (6). The detection computes the value of the output density function, under each hypothesis, at $Y_k=\eta$, mean observed over the k-th window. The greatest numerical value determines which hypothesis is valid.

The corresponding probability density function, (3) or (4), conditional upon observed output is then updated by applying the Bayesian formula:

$$ f_{I_k/Y_k}(I) = C.\ f_{I_k/Y_{k-1}}(I).\ f_V(\eta - I) $$

where C is a normalisation coefficient.

The expansion of this expression yields Lagrange multipliers λ_i, $(\ i=0,1,2)$ of the density conditional upon observed pixels:

$$ \lambda_0(k/k) = \lambda_0(k/k-1) + \mu_0 + \mu_1\eta + \mu_2\eta^2 $$
$$ \lambda_1(k/k) = \lambda_1(k/k-1) - \mu_1 - 2\mu_2\eta $$
$$ \lambda_2(k/k) = \lambda_2(k/k-1) + \mu_2 $$

where μ_i, *(i=0,1,2)* are Lagrange multipliers of the output noise density function.
The density function detected is thus updated.

Remains to predict the posterior distribution at (k+1) conditional upon the k-th observation.
The prediction uses the state equation (5) to compute the linear constraints $l_i(k+1/k)$, *(i=0,1,2)* through a transformation of random variables:

$$l_i(k+1/k) = E[\varphi_i(I_{k+1})] = E[\varphi_i(I_k, W_k)]$$

Having Gaussian approximations, analytical expressions exists. Thus,

$$\begin{aligned} l_0 &= \int f_{k+1/k}(I)dI = \iint f_{k/k}(I)f_W(w)dIdw=1 \\ l_1 &= \int I.f_{k+1/k}(I)dI = \iint ((1-a_j)I+b_j+w).f_{k/k}(I)f_W(w)dIdw \\ &= l_0((1-a_j)m_{k/k} + b_j) \\ l_2 &= \int I^2.f_{k+1/k}(I)dI = \iint ((1-a_j)I+b_j+w)^2 f_{k/k}(I)f_W(w)dIdw \\ &= l_0(q+(1-a_j)(\sigma_{k/k}^2+m_{k/k}^2)+b_j^2+2(1-a_j)b_j.m_{k/k}) \end{aligned}$$

where q is state noise variance. The values of constant coefficients, a_j and b_j, depend on the hypothesis H_j, *(j=1 or j=2)* validated.
The statistical characteristics result from linear constraints just calculated, as they are moments of the distribution.

The predicted distribution which maximises the entropy under the linear constraints is the distribution for which the Lagrange multipliers of the density function solve the non linear system:

$$l_i(k+1/k) = E[\varphi_i(I_{k+1})]]$$

Due to Gaussian approximation, the system is solved analytically. Thus, the Lagrange multipliers of the posterior density function are defined in terms of posterior mean $m_{k+1/k}$ and variance $\sigma_{k+1/k}^2$ as follows:

$$\begin{aligned} \lambda_2(k+1/k) &= -1/(2\sigma_{k+1/k}^2) \\ \lambda_1(k+1/k) &= m_{k+1/k}/\sigma_{k+1/k}^2 \\ \lambda_0(k+1/k) &= -0.5\ln(2\pi\sigma_{k+1/k}^2)+\ln(l_0) \\ &\quad - m_{k+1/k}/(2\sigma_{k+1/k}^2) \end{aligned}$$

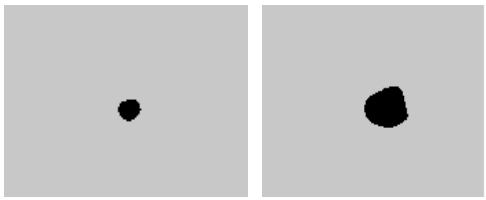
Finally, the predicted distribution conditional upon the observation window is characterised through its exponential density function with Lagrange multipliers predicted, in accordance with the FAD-concept.

5. RESULTS

This section presents the segmentation of mammograms 1 and 2, we achieved in user real time. From these two mammograms, circumscribed masses are to be located. According to a doctor, there is a small tumour in the central area of the breast 1, and there should be a bigger tumour in the top-right area of mammogram 2.

Each segmented image clearly indicates the area of the tumour obtained through the filtering by approximated densities. The respective areas, magnified two times, from segmented mammograms are displayed:

Figure 3 : Segmented images around tumours



Segmented part of mammogram 1 | Segmented part of mammogram 2

6. CONCLUSION

Through this application, the filtering by approximated densities applicability is demonstrated.

The application, using statistical characteristics up to second order, enables a precise location of tumours as circumscribed masses that can be very small or even non detectable by human eye.

As observed in section 3., some parts of mammary gland areas appears on segmented images. Further developments of FAD using higher order statistics are currently led. First results are very encouraging and we hope to reach the point where the detection/classification of other types of breast abnormalities would be possible.

REFERENCES

[1] Banks, S.P, 'Signal Processing, Image Processing and Pattern Recognition', *Prentice Hall*, 1990.

[2] de Brucq, D. & Ruiz, V., 'Modèle d'état à coefficient aléatoire. Filtrage par densités approchées', *Revue Traitement du signa'l*, Vol. 11, n°2, pp 133-140, 1994.

[3] Stathaki, T., 'Image analysis using higher order spectra with application to breast cancer localisation', *MPhil to PhD Transfer report*, Dept of Electrical and Electronic Engineering, Imperial College of London, March 1994.