Comparing Methods to Denoise Mammographic Images

P. Mayo[1], F. Rodenas[2], G. Verdú[1]
[1]Department of Nuclear and Chemical Engineering, Universidad Politécnica de Valencia, Spain
[2]Department of Applied Mathematics, Universidad Politécnica de Valencia, Spain

Abstract— **Digital mammographic image processing often requires a previous application of filters to reduce the noise level of the image while preserving important details. This may improve the quality of digital mammographic images and contribute to an accurate diagnosis. In the literature, one can find a large amount of denoising techniques available for different kinds of images. We have adapted some of the existing denoising algorithms to mammographic images. In this work, we compare the effect of different denoising filters acting on digitized mammograms. The considered filters are: a local Wiener filter, a wavelet filter, a filter based on independent component analysis, and finally, a filter based on the diffusion equation. The noise reduction is measured by the mean squared error.**

Keywords— **Digital mammography, denoising, independent component analysis, wavelet shrinkage.**

I. Introduction

The most effective technique for detecting breast occult tumours is the mammography. The low contrast of the small tumours to the background, which is sometimes close to the noise, makes that small breast cancer lesions can hardly be seen in the mammography [1]. In this sense, an image preprocessing to reduce the noise level of the image preserving the mammography structures is an important item to improve the detection of mammographic features.

Classically, denoising methods have been based on apply linear filters as the Wiener filter to the image, however linear methods tend to blur the edge structure of the image. Several denoising methods based on nonlinear filters have been introduced to avoid this problem [2,3,4].

In this work, a comparative study of several denoising techniques for mammographic images is presented. The filters considered are: 1) a local Wiener filter, 2) a filter based on the denoising method of Donoho [2] based on the minimax thresholding strategy, roughly speaking, based on a soft thresholding of the wavelet transformed coefficients of the image, and, 3) a filter based on the independent component analysis of the image [4,5,6].

II. Denoising methods

A. Adaptative Wiener filter

The classical denoising filter is the Wiener filter, defined as the linear filter that minimizes the mean squared error (MSE).

The first denoising method used in this work consists in applying a Wiener filter to an image adaptively, tailoring itself to the local image variance. Where the variance is large, the Wiener filter performs little smoothing. Where the variance is small, the Wiener filter performs more smoothing.

This approach often produces better results than linear filtering. The adaptive filter is more selective than a comparable linear filter, preserving edges and other high frequency parts of an image.

B. Wavelet shrinkage

Another denoising method is the so-called *wavelet shrinkage*, originally proposed by Donoho [2]. This method is based on the wavelet decomposition of the image, which is the two dimensional version of the wavelet decomposition of a signal.

Wavelet decomposition of a signal is a representation of the signal onto a basis of wavelet functions (see e.g. [7] for a review). Wavelets are families of basis functions, each family generated by scaling and translating a basic function called *mother wavelet*. An essential property of the wavelets is that they are well localized in time (or in the case of image, in space) and in frequency.

Typically, the (discrete) wavelet transform of a signal f is another signal of the same length, which consists in two subsignals of half its length:

$$f \rightarrow (a_1|d_1) \qquad (1)$$

The subsignal a_1 is a smoothed version of the original signal, often called approximation subsignal, and the subsignal d_1 contains high frequency information and it is often called subsignal of the details.

The key point of the wavelet denoising method is that in the wavelet domain the noise is spread fairly uniformly among all coefficients, whereas the signal is quite sparse, being concentrated into a small number of coefficients [2]. This is the practical motivation for thresholding of the detail coefficients proposed by Donoho. The threshold value T proposed by Donoho based on minimax principles, is given by [2]:

$$T = \frac{\sigma \sqrt{2 \log n}}{\sqrt{n}} \qquad (2)$$

where n is the length of the data and σ the standard deviation of the noise.

The last step of the wavelet shrinkage denoising algorithm consists in taking the inverse wavelet transform to obtain the denoised signal (or denoised image, in the 2-dimensional case).

C. ICA-based denoising method

Independent component analysis (ICA) is a method to represent a set of multidimensional data vectors in a basis where the components are as independent as possible [8,9]. Often this means that one must find a transformation that provides a vector whose components are sparse. This means that the probability of a component to be significantly different to zero is very low. For practical purposes, a random variable is called sparse when it has a supergaussian probability distribution, so it has a probability density function sharper than the gaussian density function.

ICA denoising methods rely on the fact that the transformed components have sparse (supergaussian) distributions, so that the denoising techniques attempt to reduce gaussian noise by shrinkage (soft thresholding) of these sparse components. The choice of a shrinkage function depends on the statistical distribution of each sparse component [5]. It has been shown [10] that the statistical distributions of the independent components of mammographic images are appropriated to apply the shrinkage algorithm introduced in [4,5].

In ICA an observed random vector is expressed as a linear transformation of another variables that are nongaussian and statistically independent. Denote by x the n-dimensional data vector, in our case the vector contains the pixel gray levels of an image window. The basic ICA model may be expressed as (see e.g. [8] for a comprehensive treatment):

$$x= As \tag{3}$$

where $x=[x_1,...,x_n]^T$ is the vector of observed data, $s=[s_1,...,s_m]^T$ is the vector of independent components, called source signals, and A is a constant full rank nxm matrix, named the mixing matrix. Independent components and mixing matrix are determined by requiring that the coefficients s_i, are mutually independent or as independent as possible.

The independent components are estimated by determining an mxn separating matrix W, so that the components s_i of the linear transformed vector s have maximally non-gaussian distributions and are mutually uncorrelated,

$$s= Wx \tag{4}$$

The separating matrix W is determined using an algorithm that optimizes iteratively statistical independence of the components of s. The algorithm performing ICA that we have used is the Hyvärinens *FastIca* algorithm [11].

In order to apply the ICA algorithm, the original data x must be pre-processed. First, data are centered, i.e, we subtract the data mean. The second step, called whitening, is to remove the second order statistical dependence in the data. Whitened data have unit variance and are uncorrelated. Whitening can be done using standard PCA, which simultaneously may be used to reduce the data dimension. The m-dimensional whitened vector is obtained by the linear transformation

$$y= Vx \tag{5}$$

where the mxn whitening matrix is of the form $V=D^{-1/2}E^T$. Using the eigenvalue decomposition of the covariance matrix of x, the diagonal matrix D contains the m greatest eigenvalues of the covariance matrix and the columns of the orthogonal matrix E contain the corresponding eigenvectors. After whitening, we seek an orthogonal matrix B which maximizes some given measure of supergaussianity of the components of the vector $s=By$, this may be done by using the FastICA algorithm [11]. The separating equation is then

$$s= By = BVx= BD^{-1/2}E^Tx=Wx. \tag{6}$$

If the ICA model holds, the independent components are sparse which means that each component has a supergaussian distribution. This is fundamental to apply the approach of [4,5,6] to eliminate the gaussian noise from a nongaussian random variable.

Denote by x the noisy observed random variable, the model for x can be expressed:

$$x=s+n \tag{7}$$

where s is a non-gaussian random variable corrupted by an additive gaussian noise $n\sim N(0,\sigma^2)$. The method introduced in [4,5] to denoise the observed data x proposes an estimates of s given by

$$\hat{s}=g(x) \tag{8}$$

where g is a function depending on the probability density distribution of s. The function g is a shrinkage function that can be considered a soft thresholding operator applied to the values of x.

The density distributions for natural images have been parametrized in [5,6]. Parametrizations depend on two parameters and they model different degrees of non-gaussianity.

The method of denoising a random vector consists in applying the method described above for a scalar variable to each component separately. In general, there is no guarantee that the vector components are sparse. To solve this problem, the vector is linearly transformed in such a way that the resulting components are as nongaussian as possible. Therefore, the denoising method for a vector variable consists in applying on the ICA independent components a component-wise denoising using the appropriate shrinkage function g for each component.

It is possible applying the ICA denoising method [4,5] to mammographic images [10]. The density functions analysis of the independent components of mammographic images show that these distributions are suitable to apply the ICA denoising method.

III. METHODOLOGY

Mammographic images considered in this work for denoising experiments were chosen from the MIAS MiniMammographic Database, provided by the Mammographic Image Analysis Society. The mammograms were digitised at 200 micron pixel edge, resulting images with 1024x1024 pixel resolution.

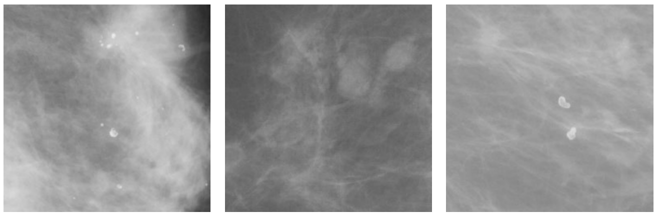

Fig. 1. Some typical mammographic images from MIAS database.

The algorithms of the denoising methods have been implemented by using MatLab 6.5 .

The adaptive Wiener filter is a local low pass filter that is processed adaptively in a local neighbourhood of 3x3 pixels blocks of the image, estimating the local image mean and standard deviation of each of them.

For the wavelet shrinkage we have chosen the coiflet 6 family at level four because of its efficient energy compactness preserving the essential information of the image.

For the ICA method, the training images to estimate the ICA transform are selected from the same MIAS MiniMammographic Database. The criterion for image selection was that the training set must be representative of the mammographic images.

Each image of the training data set was linearly normalized, so that pixels had zero mean and unit variance. The data vectors x are obtained from 20000 image patches of size 16x16 that were taken at random from the training images, so these subimages were vectorized into 256-dimensional vectors which were used as the mixed data of the ICA model. The dimension reduction is performing in such a way that the retained variance is the 98% of the initial variance of the data.

In order to estimate the sparseness of the independent components, we sample 40000 image patches at random locations from the same dataset that was used for estimation of the transform. Then we transform these samples using the estimated ICA transform and we calculate the kurtosis for each one of the components. The normalized kurtosis values encountered for every transformed component are greater than zero, this means that ICA transform find sparse representations of the original vectors. As a result, the statistics of the independent components of mammographic images are suitable for the ICA-based denoising procedure.

For a mammographic image of the database, we apply the denoising algorithm taking a sliding window approach of the image [5,6], so we apply the algorithm to all vectors obtained from every possible 16x16 window of the image. Then, each 256-dimensional vector is pre-processed by whitening and their dimension is reduced, resulting vectors are transformed into the sparse basis and the estimated nonlinearity shrinkage functions are applied to every component of each vector. After that, we invert the transformation to obtain estimates of the denoised vectors. Finally, it is necessary the reconstruction of the denoised image from the denoised vectors. Since we consider the sliding window approach, then each pixel has 256 different suggested values and we compute the final result as the mean of these values. This is a type of local filtering, considering all possible 16x16 neighbourhoods around each pixel.

IV. RESULTS

In order to perform the denoising experiments some images were selected from the MIAS database. Original images from the database were corrupted by adding a Gaussian noise. Experiments were performed using noise of standard deviation of 0.3 and 0.5 times the standard deviation of the original image. These noisy images were subsequently denoised using the adaptative Wiener filter, the Donoho wavelet shrinkage and the ICA filter.

The comparison among the different denoising results was quantitative measured by using the squared root of the mean squared error (RMSE) between the denoised image and the original noise free image (Fig.1). The RMSE is calculated by:

$$\mathrm{RMSE} = \sqrt{\frac{\sum (\mathrm{I}(x,y) - \hat{\mathrm{I}}(x,y))^2}{n \cdot m}} \tag{9}$$

where $I(x,y)$ and $\hat{I}(x,y)$ are the pixel values of the original free noise and denoised images respectively, and the size of the image is given by $n \cdot m$.

Table 1 shows the results for the RMSE obtained for two different images randomly selected from the database by using the different denoising methods. One of the images is given in Fig. 2

TABLE I
RMSE FOR THE THREE DENOISING METHODS

Images	Noisy	Wiener	Wavelet	ICA
Image1+noise 0.3σ	9.0771	3.5699	3.2201	3.7119
Image1+noise0.5 σ	15.054	5.9297	3.6083	4.0411
Image2+noise 0.3σ	4.9476	2.4494	3.3928	3.7727
Image-+noise 0.5σ	8.2874	3.3876	3.7462	4.2796

For visual evaluation of denoising methods, in Fig. 2 are shown Image 1, a noisy version of it and the resulting denoised images obtained applying the three considered methods.

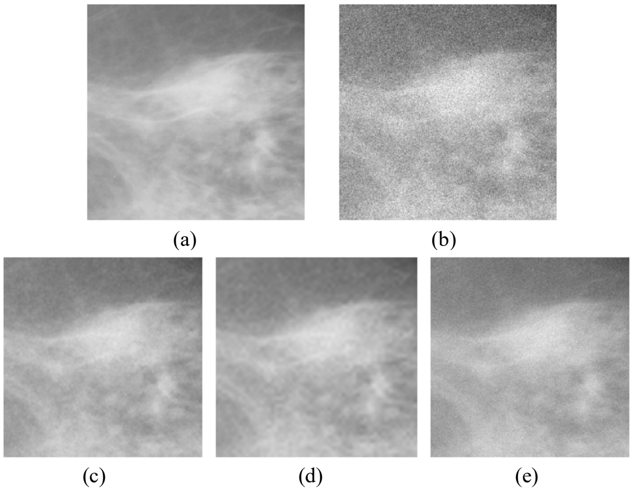

Fig. 2. Denoising process. (a) Original Image 1 of the database. (b) Noisy image, noise level added 0.5 (c) Image obtained applying Wiener filter (d) Image obtained applying wavelet filter (d) Image obtained applying ICA filter.

V. Discussion

We have used the mean squared error to quantify the succes of the three denoising methods considered above when they are applied to mammographic images. The MSE is a standard measure to evaluate a signal denoising technique, however, for image denoising, this measure it is not always related with the denoising visual results. Since the goal of the mammographic images processing is the improvement of the early breast cancer detection being the image denoising a first step of this program, then, the assessment of denoising methods would be done in the context of the use of these mammographic images.

As a result, the denoising methods compared here are suitable for application to mamographic images. The denoising results for the three cases are comparable from the MSE and visual point of view.

Acknowledgment

Financial support from "Valencian breast cancer early detection programme" and "Vicerrectorado de investigación, desarrollo e innovación of the Polytechnic University of Valencia" are gratefully acknowledged.

References

[1] J. Dengler, S. Behrens, J. F. Desaga, "Segmentation of microcalcifications in mammograms", *IEEE Trans. on Medical Imaging,* vol. 12, pp. 634-642, 1993.

[2] D. Donoho, I. Johnstone, G. Kerkyacharian, D. Picard, "Wavelet shrinkage: asymptopia?", *Journal of the Royal Statistical Society* B, vol. 57, pp. 301-369, 1995.

[3] F. Catté, P. Lions, J. Morel, T. Coll, "Image selective smoothing and edge detection by nonlinear diffusion", *SIAM Numerical Analysis*, vol. 29, pp. 182-193, 1992.

[4] A. Hyvärinen, "Sparse code shrinkage: denoising of nongaussian data by maximum likelihood estimation", *Neural Comput.* vol. 11, pp. 1739–1768, 1999.

[5] A. Hyvärinen, P. Hoyer, E. Oja, "Image denoising by sparse code shrinkage". In *Intelligent Signal Processing,* S. Hykin, B. Kosko, Eds. IEEE Press, 2001.

[6] P. Hoyer, "Independent component analysis of image denoising", Master's Thesis, Helsinki University of Technology, 1999.

[7] S. Mallat, *A wavelet tour of signal processing*. Academic Press, New York, 1998.

[8] A. Hyvärinen, J. Karhunen, E. Oja, "Independent component analysis", in *Wiley series on adaptive and learning systems for signal processing communications and control*, S. Haykin Ed. Wiley, New York, 2001.

[9] P. Comon, "Independent component analysis—a new concept?", *Signal Process*, vol. 36, pp. 287–314, 1994.

[10] P. Mayo, F. Rodenas, G. Verdú, "Denoising mammographic images using ICA", preprint of the Department of Nuclear and Chemical Engineering, Polytechnic University of Valencia (Spain), 2004, unpublished.

[11] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis", *IEEE Trans. Neural* Networks, vol. 10, pp. 626–634, 1999.