UNIVERSITÀ DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
E METODI MATEMATICI°

QUADERNI DEL DIPARTIMENTO

Department of Information Technology and Mathematical Methods

Working Paper

Series "*Mathematics and Statistics*"

n. 9/MS – 2012

Mammogram Diagnostics via 2-D Complex Wavelet-based Self-similarity Measures

by

O. Nicolis, S. Jeon, B. Vidakovic

° Viale Marconi. 5, I – 24044 Dalmine (BG), ITALY, Tel. +39-035-2052339; Fax. +39-035-562779

[§] L'accesso alle *Series* è approvato dal Comitato di Redazione. I *Working Papers* della Collana dei Quaderni del Dipartimento di Ingegneria dell'Informazione e Metodi Matematici costituiscono un servizio atto a fornire la tempestiva divulgazione dei risultati dell'attività di ricerca, siano essi in forma provvisoria o definitiva.

Mammogram Diagnostics via 2-D Complex Wavelet-based Self-similarity Measures

Orietta Nicolis[a], Seonghye Jeon[b], and Brani Vidakovic[b]

[a] *University of Bergamo (Italy),*

[b] *Georgia Institute of Technology, GA (USA),*

March 20, 2012

Abstract

Breast cancer is the second leading cause of death in women in the United States. Mammography is currently the most effective method for detecting breast cancer early; however, radiological interpretation of mammogram images remains a challenging task. On the other hand, many medical images demonstrate a certain degree of self-similarity over a range of scales which can guide us in their description and classification.

In this work, we generalize the scale-mixing wavelet transform to the complex wavelet domain. In this domain, we estimate Hurst parameter and phase and use them as discriminatory descriptors to classify mammographic images to benign and malignant. The proposed methodology is tested on a set of images from the University of South Florida Digital Database for Screening Mammography (DDSM).

Key words: Scaling; Self-similarity; Complex Wavelets; Image Classification

1 Introduction

The National Cancer Institute estimates that 1 in 8 women born today will be diagnosed with breast cancer during her lifetime (Altekruse *et al.*, 2010). Breast cancer is one of the most common forms of cancer among women in

the United States, second only to non-melanoma skin cancer. A national objective has been set by the U.S. Department of Health and Human Services to reduce the female breast cancer death rate from 22.9 per 100,000 females in 2007 down to 20.6 by the year 2020 – a 10% improvement (Healthy People 2020, U.S. Department of Health and Human Services). One of the most important tools toward that goal is advanced precision of screening technologies. Early detection is the best method for improving prognosis and also leads to less invasive options for both specific diagnosis and treatment.

Mammography is currently the most efficient and prevalent method for detecting a breast cancer early, before it is substantial enough to cause symptoms. However, the radiological interpretation of mammogram images is a difficult task since the appearance of even normal tissue is highly variable and complex, and signs of early disease are often small or indistinct. Suspicious findings are commonly clarified by follow-up images, ultrasound, or MRI. On the other hand, it has been estimated that $10 - 30\%$ of cancers which could have been detected are missed (Martin *et al.*, 1979). Thus, improving both the specificity and the sensitivity of mammographic diagnoses is an important goal in improving prognoses while also reducing the number of unnecessary procedures or surgical operations.

In high frequency and irregular data collected in real-life settings (both naturally occurring and human-made), a commonly occurring phenomenon is that of regular scaling. Examples of this have been found in a variety of systems and processes including economics (stock market, exchange rate fluctuations), telecommunications (internet data), physics (hydrology, turbulence), geosciences (wind and rainfall patterns), and several applications in biology and medicine (DNA sequences, heart rate variability, auditory nerve-spike trains). The irregular behaviors of these complex structures are difficult or impossible to quantify by standard modeling techniques. But when observations are inspected at different scales, there is in fact a regular relationship between the behavior and the scale. This phenomenon has been demonstrated in many medical images, leading to diagnostic use of tools capable of quantifying statistical similarity of data patterns at various scales.

The standard measure of regular scaling is the Hurst exponent. This measure can also be connected to phenomena of long memory and fractality in signals and images and is viewed as an informative summary. Many techniques for estimating the Hurst exponent exist, and assessing the accuracy of these estimations can be complicated. Wavelet transforms are powerful tool in estimating the Hurst exponent and modeling statistical similarity

at different scales. For example, Nicolis *et al.* (2011) proposed a method based on the wavelet spectra for extracting the self-similarity measures in an isotropic and anisotropic spaces. Ramírez-Cobo *et al.* (2011) demonstrated a wavelet-based spectra method for estimating Hurst exponent in time-varying two-dimensional rainfall maps.

For an efficient representation of the image or signal, the wavelet basis is desirable to be orthogonal, symmetric and to have compact support. (Gao and Yan, 2010). An orthogonal basis has a variety of theoretical and practical advantages: it leads to more efficient algorithms, and establishing properties of a representation is often easier with orthogonal bases. Symmetry guarantees an orientation-free representation of features, preventing distortion in the data induced by its basis representation. Moreover, the computational cost of performing wavelet transforms depends heavily on the support size of a basis. Apart from the Haar wavelet, complex wavelets with an odd number of vanishing moments are only compactly supported wavelets which are symmetric (Lawton, 1993). Due to this advantages, complex wavelet has been used in various areas including motion estimation (Magarey and Kingsbury, 1998), texture image modeling (Portilla and Simoncelli, 2000), image denoising (Achim and Kuruoglu, 2005) and NMR spectra classification (Kim *et al.*, 2008).

The novelty of this paper is to use the scale-mixing wavelet spectra based on *complex wavelet transforms* for estimating the Hurst exponent. We then focus on the estimated Hurst exponent and show its ability to differentiate cancerous from normal tissue visible in the backgrounds of mammogram images, and compare this performance with its counterpart obtained from real-valued wavelet transform. Moreover complex wavelet transform produces an additional measure, phase information. We also demonstrate the classification power of the phase information and use it as an additional modality in the discriminatory analysis.

A further novelty of our work is the use of the information contained in the background tissue of images. Most of the references found in literature dealing with breast cancer detection methods are based on microcalcifications (Wang and Karayiannis, 1998; Netsch and Peitgen, 1999; Kestener *et al.*, 2001; El-Naqa *et al.*, 2002). Only recently the information contained in the background is taken into consideration (Nicolis *et al.*, 2011; Hamilton *et al.*, 2011). This classifying measure based on background tissue would be a new tool to be used in combination with existing clinical diagnostic tools, thus improving the power of non-invasive diagnostic techniques.

The paper is organized as follows. In Section 2, we briefly describe the data set used in the analysis. In Section 3, the complex wavelet based scale-mixing wavelet spectra is proposed as new tools for estimating the self-similarity indices. Also the importance of the phase information is discussed. Section 4 deals with the classification of mammogram images using the wavelet descriptors obtained from Section 3 as classifiers. Finally Section 5 contains concluding remarks.

2 The Dataset

The collection of digitized mammograms we analyzed was obtained from the University of South Florida's Digital Database for Screening Mammography (DDSM)

`http://marathon.csee.usf.edu/Mammography/Database.html`.

The DDSM is described in detail in Heath *et al.* (2000). Images from this database containing suspicious areas are accompanied by pixel-level "ground truth" information relating locations of suspicious regions to what was assessed and verified through biopsy. We selected 105 normal (benign) cases from volumes normal-01, and 98 cancer cases from volumes cancer-01 and cancer-02. Each case contains four mammograms (two for each breast: the craniocaudal (CC) and mediolateral oblique (MLO) projections) from a screening exam. We considered only the CC projections, using the right breast image for all normal cases, and the cancerous breast (right or left) image for cancer cases. A sub-image of size 1024×1024 was taken from each case for analysis. An example of sub-image is provided in Fig. 1.

3 Discrete complex wavelets

This section discusses discrete complex wavelet transforms. Unlike the popular method used in this context (Selesnick *et al.*, 2005), the proposed method is orthogonal and most parsimonious.

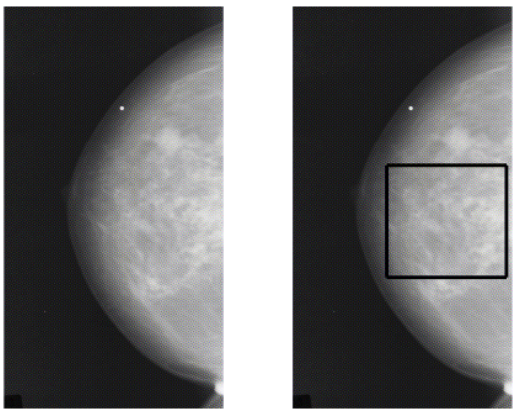


Figure 1: *Left panel*: right CC mammogram corresponding to a malignant case. *Right panel*: sub-image of size 1024×1024 considered for the analysis.

3.1 Complex wavelet basis

The construction of the complex wavelet basis associated with multiresolution analysis follows the usual approach proposed by Mallat (1998) and Daubechies (1992). Details on the construction and properties of complex wavelets can be found in Lawton (1993), Lina and Mayrand (1995), Strang and Nguyen (1996), and Zhang *et al.* (1999). In analogy to the real case, the wavelet function $\psi(x)$ for the complex wavelet is given by

$$\psi(x) = \frac{1}{\sqrt{2}} \sum_k (-1)^k 2h^*_{1-k} \phi(2x - k) \tag{1}$$

where ϕ is the scaling function, h is the low pass filter and the * indicates the complex conjugate. The representation of wavelets in 2-D can be done through the tensor product of univariate scaling functions and wavelets as follows:

$$\begin{aligned} \phi(x_1, x_2) &= \phi(x_1) \cdot \phi(x_2) \\ \psi_h(x_1, x_2) &= \phi(x_1) \cdot \psi(x_2) \\ \psi_v(x_1, x_2) &= \psi(x_1) \cdot \phi(x_2) \\ \psi_d(x_1, x_2) &= \psi(x_1) \cdot \psi(x_2) \end{aligned} \tag{2}$$

where symbols h, v, d in (2) stand for horizontal, vertical and diagonal directions, respectively. The atoms capture image features in the corresponding directions.

3.2 The complex scale-mixing 2-D wavelet transform

The discrete complex wavelet transform (DCWT) can be considered as a complex-valued extension of the standard discrete wavelet transform (DWT). It uses complex-valued filtering (analytic filter) for transforming the real/complex signals. Complex wavelet coefficients can be computed by Mallat's algorithm (Mallat, 1998)

$$c_{j-1,l} = \sum_k h^*_{k-2l} c_{j,k} \tag{3}$$

and

$$d_{j-1,l} = \sum_k g^*_{k-2l} c_{j,k} \tag{4}$$

where h is as in (1) and g is the quadrature mirror filter. The * denotes the complex conjugate. Conversely, the reconstruction is given by

$$c_{j,k} = \sum_l c_{j-1,l} h_{k-2l} + \sum_l d_{j-1,l} g_{k-2l}. \tag{5}$$

The real and imaginary coefficients are used to compute the modulus and phase information. The wavelet coefficients can be written as

$$d_{j,k} = Re(d_{j,k}) + i \cdot Im(d_{j,k})$$

with magnitude

$$|d_{j,k}| = \sqrt{Re(d_{j,k})^2 + Im(d_{j,k})^2}$$

and phase

$$\angle d_{j,k} = \arctan\left(\frac{Im(d_{j,k})}{Re(d_{j,k})}\right)$$

when $|Re(d_{j,k})| > 0$.

There are many versions of 2-D wavelet transforms which lead to different tessellations, or tilings (Ramírez-Cobo *et al.*, 2011). We define the complex wavelet atoms as follows

$$\begin{aligned} \phi_{(j_1,j_2),\mathbf{k}}(\mathbf{x}) &= 2^{(j_1+j_2)/2}\ \phi(2^{j_1}x_1-k_1, 2^{j_2}x_2-k_2) \quad (6)\\ \psi_{\delta,(j_1,j_2),\mathbf{k}}(\mathbf{x}) &= 2^{(j_1+j_2)/2}\ \psi_{\delta}(2^{j_1}x_1-k_1, 2^{j_2}x_2-k_2), \quad (7) \end{aligned}$$

where δ is one of directions h, v, or d, and $(j_1, j_2), (k_1, k_2) \in \mathbb{Z}^2$. Then, any function $f \in \mathcal{L}_2(\mathbb{R}^2)$ can be represented as

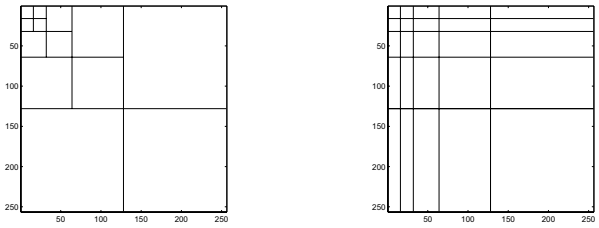
$$\begin{aligned} f(\mathbf{x}) &= \sum_{\mathbf{k}} c_{(J_0,J_0),\mathbf{k}}\ \phi_{(J_0,J_0),\mathbf{k}}(\mathbf{x}) \\ &+ \sum_{j>J_0}\sum_{\mathbf{k}} d_{(J_0,j),\mathbf{k}}\ \psi_{h,(J_0,j),\mathbf{k}}(\mathbf{x}) \\ &+ \sum_{j>J_0}\sum_{\mathbf{k}} d_{(j,J_0),\mathbf{k}}\ \psi_{v,(j,J_0),\mathbf{k}}(\mathbf{x}) \\ &+ \sum_{j_1,j_2>J_0}\sum_{\mathbf{k}} d_{(j_1,j_2),\mathbf{k}}\ \psi_{d,(j_1,j_2),\mathbf{k}}(\mathbf{x}), \end{aligned}$$

and a 2-D wavelet transform, which we call the *scale-mixing wavelet transform*, is obtained. The scale-mixing detail coefficients are defined as

$$\begin{aligned} d_{(J_0,j),\mathbf{k}} &= \int f(\mathbf{x})\ \psi^*_{h,(J_0,j),\mathbf{k}}(\mathbf{x})d\mathbf{x}, \\ d_{(j,J_0),\mathbf{k}} &= \int f(\mathbf{x})\ \psi^*_{v,(j,J_0),\mathbf{k}}(\mathbf{x})d\mathbf{x}, \\ d_{(j_1,j_2),\mathbf{k}} &= \int f(\mathbf{x})\ \psi^*_{d,(j_1,j_2),\mathbf{k}}(\mathbf{x})d\mathbf{x}, \quad (8) \end{aligned}$$

where ψ^* is a complex conjugate of ψ. Note that (j_1, j_2) in (6) and (7) can be indexed as $(j_1, j_1 + s)$ with $s \in \mathbb{Z}$.

Similarly to the traditional one- and two-dimensional cases, the complex scale-mixing detail coefficients are linked to the original image (2-D signal) through a matrix equation. Suppose that a $2^n \times 2^n$ image (matrix) A is to be transformed into the wavelet domain. The complex wavelet matrix W is first composed by the complex scaling and wavelet filter coefficients h_k and g_k as in Vidakovic (1999). Note that the wavelet filter is given by $g_k = (-1)^k h^*_{1+N-k}$ and N is a shift parameter which affects the location of the wavelet. Then



(a) (b)

Figure 2: Tessellations for 2-D wavelet transforms. (a) Traditional 2-D transform of depth 4; (b) Scale-mixing wavelet transform of depth 4.

the rows of A are transformed by a one-dimensional transform given by the wavelet matrix W, resulting in WA'. The same is repeated on the rows of WA'. The result is

$$B = W(WA')' = WAW', \tag{9}$$

the scale-mixing wavelet transform of matrix A, which will be the basis for defining the scale-mixing spectra. It represents a finite-dimensional implementation of (8) for signal $f(\mathbf{x})$ sampled in a form of matrix A.

The tessellation induced by transform in (9) is shown in Figure 2 (b). A more general transform can be obtained as an iterative repetition of the transform in (9) with depth k, applied only on the "smooth part" of the previous iterative step.

The scale-mixing 2-D transform is operationally appealing. Constructing appropriate W is computationally fast and, since W is orthogonal, the inverse transform is straightforward:

$$A = W'BW.$$

By inspecting the tessellation in Figure 2, several hierarchies of detail spaces can be identified. The diagonal hierarchy interfaces coefficients with the same component scales and coincides with the diagonal hierarchy in the traditional 2-D spectra. One level above and below the diagonal hierarchy are hierarchies of detail spaces that interface the scales that differ by 1. For the hierarchy above the diagonal, the scales along x_1-direction are interfaced by the next coarser scale along x_2-direction. For the hierarchy below the diagonal, the roles of x_1 and x_2 are interchanged.

The orthogonality of W implies

$$\text{trace}(AA') = \text{trace}(BB')$$

for $B = WAW'$, implying the total energy in the image A

$$E = \text{trace}(AA')$$

is preserved.

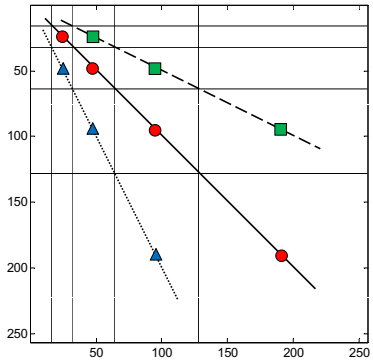


Figure 3: (a) Three detail-space hierarchies generating the scale-mixing 2-D transform, where (j_1, j_2) is indexed as $(j, j+s)$, $s \in \mathbb{Z}$. Circles correspond to $s = 0$, triangles to $s = 1$, and squares to $s = -1$. The scales (j_0, j), $j_0 = 7$ (squares), and (j, j_0), $j_0 = 6$ (triangles) are shown in the figure.

3.3 The complex scale-mixing wavelet spectra

The scale-mixing spectra is defined in terms of the complex scale-mixing coefficients in (8)

$$S(j) = \log_2 \mathbb{E}\left(|d_{(j,j+s),\mathbf{k}}|^2\right), \tag{10}$$

where $j, s \in \mathbb{Z}$ are fixed. Note that $s = 0$ in (8) corresponds to the diagonal 2-D spectra.

To calibrate the scale-mixing spectra, consider now a 2-D fractional Brownian motion, $B_H(\mathbf{u})$. For such a process, the scale-mixing detail coefficients are given by

$$d_{(j,j+s);\mathbf{k}} = \int B_H(\mathbf{u}) \psi^*_{d,(j,j+s),\mathbf{k}}(\mathbf{u}) d\mathbf{u} \tag{11}$$

where ψ_d^* denotes the complex conjugate of ψ_d, the wavelet atom in the diagonal direction defined in (7). These coefficients are random variables with zero mean and variance (Heneghan *et al.*, 1996), which leads to

$$\mathbb{E}\left[|d_{(j,j+s);\mathbf{k}}|^2\right] = 2^{2j+s} \int \psi\left(2^j u_1 - k_1, 2^{j+s} u_2 - k_2\right) \times \psi^*\left(2^j v_1 - k_1, 2^{j+s} v_2 - k_2\right) \mathbb{E}\left[B_H(\mathbf{u}) B_H(\mathbf{v})\right] d\mathbf{u}\, d\mathbf{v}. \quad (12)$$

As in Veitch and Abry (1999), we assume that the coefficients within and across the scales are uncorrelated.

From (12), it can be shown that

$$\mathbb{E}\left[|d_{(j,j+s);\mathbf{k}}|^2\right] = 2^{-j(2H+2)}\, V_{\psi,s}(H) \quad (13)$$

where $V_{\psi,s}(H)$ is an expression depending on ψ, H, and s, but not on the scale j,

$$V_{\psi,s}(H) = -\frac{\sigma_H^2}{2} \int\int \psi(\mathbf{p}+\mathbf{q})\, \psi^*(\mathbf{q})\, |\mathbf{p}_s|^{2H} 2^{-s}\, d\mathbf{p}\, d\mathbf{q}. \quad (14)$$

A proof of (13) is provided in the Appendix.

By taking logarithms in (13), we obtain

$$\log_2 \mathbb{E}\left[|d_{(j,j+s);\mathbf{k}}|^2\right] = -(2H+2)j + \log_2 V_{\psi,s}(H) \quad (15)$$

for $j \in \mathbb{Z}$. The Hurst exponent can be estimated from the slope of the linear equation (15). Finally, the empirical counterpart of (15) is a regression defined on pairs

$$\left(j, \log_2 \frac{1}{n} \sum_{j,j+s} \left|d_{(j,j+s),\mathbf{k}}\right|^2\right), \quad j, s \in \mathbb{Z}. \quad (16)$$

The slope of the regression would estimate the Hurst exponent, i.e., $H = -(slope + 2)/2$. Instead of the sample mean in (16), different location measures could be used, such as the median.

Although (16) is based on the ordinary least squares (OLS) regression, the variance of wavelet coefficients is not constant. Therefore we can improve the estimator by using more robust approach that incorporates heteroscedasticity. Veitch and Abry (1999) used weighted linear regression to improve the

estimator. This method weights each level by the inverse of the variance of that level. Hamilton *et al.* (2011) proposed estimation methods that are based on a weighted average of all pairwise slopes s_{ij} between levels i and j. Given a weight w_{ij}, the estimator of the overall slope in (16) is then $\sum_{i,j} w_{ij} s_{ij} / \sum_{i,j} w_{ij}$. Different types of weights are proposed, from which we obtain more robust estimation methods. In this paper we adopted robust estimators proposed in the literature along with the OLS regression based estimator to perform more comprehensive comparison.

3.4 The complex phase information

It is known that phase and spectrum are interacting in a nontrivial way to describe the data. While phases encode most of the coherent (in space and scale) structure of the image, the spectrum mostly encode the strength of local information that could be corrupted with noise (Clonda *et al.*, 2004). For this reason phase information have been used in edge detection and in the reconstruction of images. A classical illustration is given in Oppenheim and Lim (1981) where the image reconstruction is more driven by the phase of the Fourier transform rather than by the magnitude.

Recently, many research studies have focused on using phase information from the complex wavelet transforms (Anderson *et al.*, 2005; Hua and Orchard, 2008; Miller and Kingsbury, 2008; Rakvongthai and Oraintara, 2008). In the wavelet domain the phase of a coefficient near an isolated feature varies linearly with its distance from the feature. Despite of the numerous literatures focused on the usage of phase in detecting edges and the feature orientations, the discriminatory power of phase in the complex wavelet domain has not yet been studied and is unknown.

Although the phase of coefficients at each level does not have any scaling property as the wavelet-based spectra, the summary statistics of the phases turn out to be discriminatory. In the following section, we demonstrate how the phase information can be used as a classification modality.

4 Mammogram Classification

In this section we illustrate how the complex wavelet-based spectra and the phase information can be used to classify digitized mammograms. We demonstrate that the spectra slope and phase, as descriptors of digitized images,

have good discriminatory power. We note that it is straightforward to implement the described analysis in various scientific areas in which 2-D data are instrumental, such as geoscience or industrial applications.

For every sub-image of size 1024×1024, we performed discrete real-valued wavelet transform (DWT) and discrete complex wavelet transform (DCWT) using Daubechies 6 tap real and complex filters. Both transforms are based on the scale-mixing 2-D wavelet transform described in Section 3.2. Next, we estimated the slope of wavelet spectra using traditional ordinary least squares regression (OLS) along with four robust estimation methods. The robust approaches include Abry-Veitch weighted regression (AV), modified level enhanced OLS (MEOLS), harmonic average weighted slopes (HA) and modified HA (MHA). For more details on these robust estimators, we refer the reader to Veitch and Abry (1999); Hamilton *et al.* (2011). Note that the wavelet spectra slope is used as a predictor instead of the Hurst exponent. This is because the estimated Hurst exponents H are empirical, and the slowly decaying spectra could cause H to be negative.

For each classification method, we randomly selected 67% of the data as a training set to fit the classifier and used the remaining 33% of the data to test performance. The random selection of training and testing sets was repeated 10,000 times, so the reported prediction errors are averaged over 10,000 runs. Performance of each model was compared in terms of sensitivity, specificity, and overall correct classification rate.

The most parsimonious classification approach would be the logistic regression involving only the wavelet spectra slope as a predictor. The result is summarized in Table 1. The robust estimation methods show superior performance over OLS estimator, with rates ranging from 0.58 to 0.86. Although the performance of DWT and DCWT is comparable, overall sensitivity and correct classification rates are higher with complex wavelet transform.

One of the interesting findings is that the phase contains information to classify normal and malignant images. Since the features and directions of background tissue is best preserved in the level of finest detail, we focus on the phase information of finest detail. Figure 4 shows the estimated density of the phase average and variance at the finest level. While the average of malignant and normal cases are similar, the variance is quite different; the phase from normal images have higher variance, which implies more irregularity. This finding is consistent with universal paradigm in medical signal and image processing, that increased regularity of signals and images is often associated with pathologies. In this case, we hypothesize control mammograms have no

Table 1: Logistic classification based on the wavelet spectra slope; five different estimation methods were compared, each with real-valued (DWT) and complex (DCWT) wavelet transforms.

	Method	Sensitivity	Specificity	Correct Classification
OLS	DWT	0.2472	0.7304	0.4744
	DCWT	0.2912	0.7135	0.4892
AV	DWT	0.7396	0.7901	0.7634
	DCWT	0.7635	0.8064	0.7838
MEOLS	DWT	0.5795	0.6768	0.6248
	DCWT	0.6175	0.7039	0.6585
HA	DWT	0.6168	0.7175	0.6651
	DCWT	0.6573	0.7361	0.6958
MHA	DWT	0.8599	0.8298	0.8436
	DCWT	0.8627	0.8489	0.8545

clusters of consistent features and edges in the detailed wavelet space.

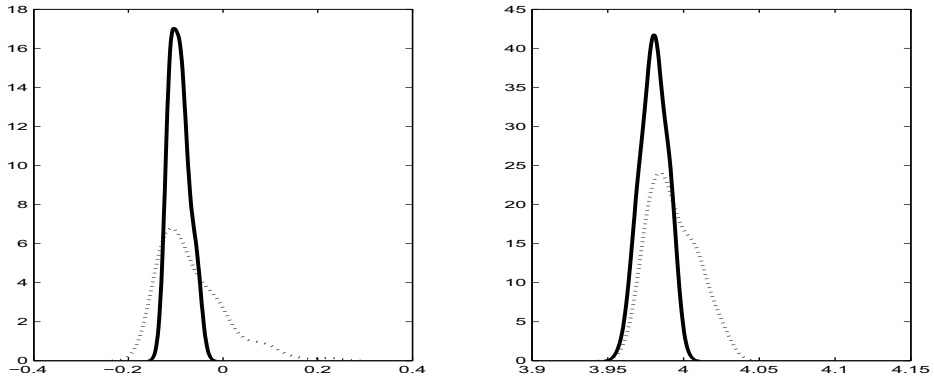


Figure 4: *Left panel*: Estimated density of phase average at the finest level. *Right panel*: Estimated density of phase variance at the finest level. Both are obtained from 105 normal cases and 98 cancer cases. The solid line corresponds to malignant cases and the dotted line to normal cases.

To validate the discriminatory power of the phase information, we assessed the logistic models by using each of the three summary statistics (average, variance, interquartile range) of the phases at the finest level. As Table 2 shows, the variance as a single predictor best classifies malignant and normal images with correct classification rate being 67%. Other measures such as skewness, or quartiles Q1 and Q3 retain classification power in the range of 50-60%.

Table 2: Logistic classification based on summary statistics of phase information at the finest level. Q3-Q1 indicates the interquartile range.

	Method	Sensitivity	Specificity	Correct Classification
Phase	Average	0.6018	0.5275	0.5533
	Variance	0.7004	0.6384	0.6654
	Q3-Q1	0.4476	0.6727	0.5520

Next, the classification analysis based on the complex wavelet spectra slope and the phase variance is conducted. Figure 5 shows a scatter plot of cases by complex spectra slope versus phase variance, illustrating the differentiation between benign and malignant cases.

We combined complex wavelet spectra obtained from five different estimation methods with the phase variance. For each of the five pairs, we performed logistic, linear, quadratic and support vector machine (SVM) classification. The result is summarized in Table 3. By comparing Table 3 with Tables 1 and 2, we conclude that the overall performance improved significantly regardless of the pair combination. The improvement is especially notable for OLS; from Table 1, the OLS estimator itself had almost no discriminatory power (correct classification rate below 0.5). However by combining it with the phase variance, the rates have increased up to 67%. Several wavelet bases and level combinations have been compared but the results and the conclusion remain the same.

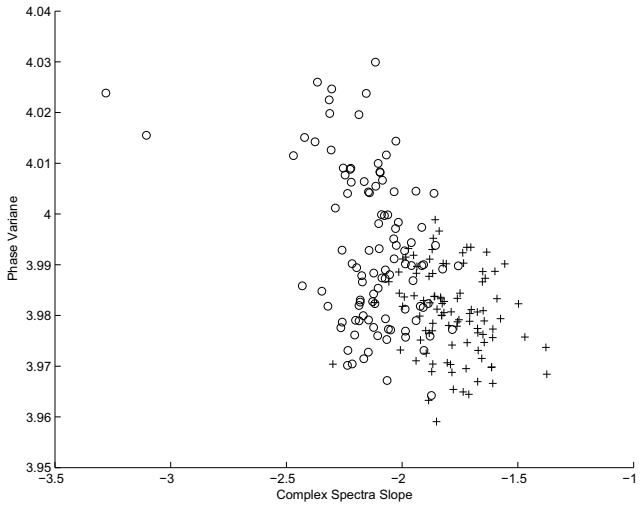


Figure 5: Scatter plot of Complex spectra slope (obtained by MHA estimation method) versus Phase variance. The symbols denote: *circles* for normal mammographies, *crosses* for malignant mammographies.

5 Conclusions

In this paper we propose a complex scale-mixing 2-D wavelet transform in the context of assessing regularities of 2-D objects. The proposed transform is implemented by matrix multiplication, and it guarantees orthogonality, directional insights, interplay between scales, and a straightforward inverse transform. We then explore the spectra and self-similarity measures based on the proposed complex wavelet transform. Their discriminatory power is demonstrated in the context of mammogram image classification. The procedure is based on background tissues of images rather than mammogram features such as microcalcifications and tumor masses, which is unused diagnostic modality in the field. The estimated Hurst exponent and phase information turn out to be discriminatory summaries in mammogram image classification. Although phase information has been used mainly for edge detection and image reconstructions in the literature, we identify that the summary measures of phase contribute to the correct classification of cancer and normal images.

Table 3: Logistic, Linear, Quadratic and SVM classification based on the pair of complex wavelet spectra slope and the phase variance.

	Method	Sensitivity	Specificity	Correct Classification
(OLS; phase variance)	Logistic	0.6900	0.6338	0.6578
	Linear	0.7376	0.6067	0.6701
	Quadratic	0.8091	0.5365	0.6687
	SVM	0.4568	0.9012	0.6723
(AV; phase variance)	Logistic	0.7851	0.8009	0.7919
	Linear	0.8171	0.7744	0.7951
	Quadratic	0.8390	0.7522	0.7943
	SVM	0.7650	0.8127	0.7882
(MEOLS; phase variance)	Logistic	0.7095	0.7265	0.7155
	Linear	0.7629	0.6548	0.7072
	Quadratic	0.8145	0.5630	0.6849
	SVM	0.5472	0.8291	0.6839
(HA; phase variance)	Logistic	0.7395	0.7589	0.7471
	Linear	0.7805	0.7003	0.7392
	Quadratic	0.8221	0.6155	0.7157
	SVM	0.6372	0.7900	0.7113
(MHA; phase variance)	Logistic	0.8561	0.8574	0.8559
	Linear	0.8814	0.8320	0.8559
	Quadratic	0.8827	0.8297	0.8554
	SVM	0.8243	0.8826	0.8525

To obtain the estimates of Hurst exponent, four robust estimation methods (AV, MEOLS, HA, MHA) along with the ordinary least squares estimator are used. From the logistic classification model, we found that robust Hurst exponent estimates and the phase variance at the finest level can differentiate between benign and malignant cases with correct classification rates 63% and above.

It is well known that for the real wavelets there is no symmetric and compactly supported scaling function defining an orthogonal MRA except for the Haar wavelet. Complex wavelets assure symmetry, compact support and orthogonality of decomposing atoms, which are desirable properties in image

representations. This advantage led to the higher sensitivity and correct classification rate across different representation scenarios.

Another benefit of using the proposed model is that we can combine the Hurst exponent estimates with the phase variance to identify whether the images contain the evidence of malignancy. By combining these two measures, we can improve correct classification rates. Note that the most accurate classification rate achieves 86% with (MHA; phase variance) pair. However, MHA and MEOLS employ weights that heavily emphasize the fine detail levels, which is in accordance with the empirical observation that finer levels in real wavelet decompositions are critical for correct mammogram image classification (Hamilton *et al.*, 2011). Therefore complex wavelets provide more accurate tool in the sense that the under-performing summary in real wavelets becomes discriminatory when combined with a phase information.

Appendix

Derivation of expression (13)

The scale-mixing detail coefficients of a 2-D fBm (11) are

$$d_{(j,j+s);\mathbf{k}} = 2^{j+\frac{s}{2}} \int B_H(\mathbf{u}) \psi \left(2^j u_1 - k_1, 2^{j+s} u_2 - k_2\right) d\mathbf{u}.$$

These coefficients are random variables with zero mean and variance (Heneghan *et al.*, 1996). Therefore,

$$\begin{aligned} \mathbb{E}\left[|d_{(j,j+s);\mathbf{k}}|^2\right] = 2^{2j+s} \int & \psi \left(2^j u_1 - k_1, 2^{j+s} u_2 - k_2\right) \\ & \times \psi^* \left(2^j v_1 - k_1, 2^{j+s} v_2 - k_2\right) \mathbb{E}\left[B_H(\mathbf{u}) B_H(\mathbf{v})\right] d\mathbf{u}\ d\mathbf{v}. \quad (17) \end{aligned}$$

Since

$$\mathbb{E}\left[B_H(\mathbf{u}) B_H(\mathbf{v})\right] = \frac{\sigma_H^2}{2} \left(|\mathbf{u}|^{2H} + |\mathbf{v}|^{2H} - |\mathbf{u} - \mathbf{v}|^{2H}\right),$$

and

$$\int \psi \left(2^j u_1 - k_1, 2^{j+s} u_2 - k_2\right) d\mathbf{u} = \int \psi \left(2^j v_1 - k_1, 2^{j+s} v_2 - k_2\right) d\mathbf{v} = 0,$$

it can be easily seen that (17) becomes

$$\begin{aligned} \mathbb{E}\left[|d_{(j,j+s);\mathbf{k}}|^2\right] = -\frac{\sigma_H^2}{2}\ 2^{2j+s} \int \int & \psi \left(2^j u_1 - k_1, 2^{j+s} u_2 - k_2\right) \\ & \times \psi^* \left(2^j v_1 - k_1, 2^{j+s} v_2 - k_2\right) |\mathbf{u} - \mathbf{v}|^{2H} d\mathbf{u}\ d\mathbf{v}. \end{aligned}$$

Next, let us define substitutions

$$\mathbf{p} \ = \ (p_1, p_2) = \left(2^j (u_1 - v_1), 2^{j+s} (u_2 - v_2)\right),$$

Then, if $\mathbf{p}_s \equiv (p_1, 2^{-s} p_2)$,

$$\begin{aligned} \mathbb{E}\left[|d_{(j,j+s);\mathbf{k}}|^2\right] &= -\frac{\sigma_H^2}{2}\ 2^{2j+s} \int \int \psi(\mathbf{p} + \mathbf{q})\ \psi^*(\mathbf{q})\ 2^{-2jH}\ |\mathbf{p}_s|^{2H}\ 2^{-4j-2s}\ d\mathbf{p}\ d\mathbf{q} \\ &= -\frac{\sigma_H^2}{2}\ 2^{-j(2H+2)} \int \int \psi(\mathbf{p} + \mathbf{q})\ \psi^*(\mathbf{q})\ |\mathbf{p}_s|^{2H} 2^{-s}\ d\mathbf{p}\ d\mathbf{q} \\ &= 2^{-j(2H+2)}\ V_{\psi,s}(H), \end{aligned}$$

where

$$V_{\psi,s}(H) = -\frac{\sigma_H^2}{2} \int\int \psi(\mathbf{p}+\mathbf{q})\ \psi^*(\mathbf{q})\ |\mathbf{p}_s|^{2H} 2^{-s}\ d\mathbf{p}\ d\mathbf{q}, \tag{18}$$

is an integral depending on ψ, H, and s, but not on the scale j.

References

Achim, A. and Kuruoglu, E. (2005). Image denoising using bivariate α-stable distributions in the complex wavelet domain. *IEEE Signal Processing Letters*, **12**, 17–20.

Altekruse, S., Kosary, C., Krapcho, M., Neyman, N., Aminou, R., and Waldron, W. (2010). Seer cancer statistics review: 1975-2007.

Anderson, R., Kingsbury, N., and Fauqueur, J. (2005). Determining multiscale image feature angles from complex wavelet phases. *International conference on Image processing (ICIP), Toronto, Canada, September 28-30*.

Clonda, D., Lina, J.-M., and Goulard, B. (2004). Complex daubechies wavelets: properties and statistical image modelling. *Signal Processing*, **84**, 1–23.

Daubechies, I. (1992). *Ten Lectures on Wavelets*. Number 61 in CBMS-NSF Series in Applied Mathematics. SIAM, Philadelphia.

El-Naqa, I., Yang, Y., Wernick, M., Galatsanos, N., and Nishikawa, R. (2002). A support vector machine approach for detection of microcalcifications. *IEEE Transactions on medical imaging*, **21**(12), 1552–1563.

Gao, R. and Yan, R. (2010). *Wavelets: Theory and Applications for Manufacturing*. Springer.

Hamilton, E. K., Jeon, S., Cobo, P. R., Lee, K. S., and Vidakovic, B. (2011). Diagnostic classification of digital mammograms by wavelet-based spectral tools: A comparative study. *The Proceedings of the 2011 IEEE International Conference on Bioinformatics and Biomedicine*, pages 384–389.

Heath, M., Bowyer, K., Kopans, D., Moore, R., and Kegelmeyer, P. (2000). The digital database for screening mammography. *The Proceedings of the 5th International Workshop on Digital Mammography, Toronto, Canada, Medical Physics Publishing (Madison, WI)*.

Heneghan, C., Lown, S., and Teich, M. (1996). Two dimensional fractional Brownian motion: Wavelet analysis and synthesis. *Image analysis and*

interpretation, proceedings of the IEEE Southwest Symposium, pages 213–217.

Hua, G. and Orchard, M. T. (2008). Image reconstruction from the phase or magnitude of its complex wavelet transform. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3261–3264.

Kestener, P., Lina, J., Saint-Jean, P., and Arneodo, A. (2001). Wavelet-based multifractal formalism to assist in diagnosis in digitized mammograms. *Image analysis and stereology*, **20**(3), 169–175.

Kim, S., Wang, Z., Oraintarac, S., Temiyasathita, C., and Wongsawatc, Y. (2008). Feature selection and classification of high-resolution nmr spectra in the complex wavelet transform domain. *Chemometrics and Intelligent Laboratory Systems*, **90**, 161–168.

Lawton, W. (1993). Applications of complex valued wavelet transforms to subband decomposition. *IEEE Transactions on Signal Processing*, **41**, 3566–3568.

Lina, J.-M. and Mayrand, M. (1995). Complex daubechies wavelets. *Applied and Computational Harmonic Analysis*, **2**, 219–229.

Magarey, J. and Kingsbury, N. (1998). Motion estimation using a complex-valued wavelet transform. *IEEE Transactions on Signal Processing*, **46**, 1069–1084.

Mallat, S. G. (1998). *A Wavelet Tour of Signal Processing*. Academic Press, San Diego.

Martin, J., Moskowitz, M., and Milbrath, J. (1979). Breast cancer missed by mammography. *American Journal of Roentgenology*, **37**(2), 142–162.

Miller, M. and Kingsbury, N. (2008). Image modeling using interscale phase properties of complex wavelet coefficients. *IEEE Transactions on Image Processing*, **17**(9), 1491–1499.

Netsch, T. and Peitgen, H. (1999). Scale-space signatures for the detection of clustered microcalcifications in digital mammograms. *IEEE Transactions on medical imaging*, **18**(9), 774–786.

Nicolis, O., Ramírez-Cobo, P., and Vidakovic, B. (2011). 2-D wavelet-based spectra with applications. *Computational Statistics and Data Analysis*, **55**(1), 738–751.

Oppenheim, A. and Lim, J. (1981). The importance of phase in signals. *The Proceedings of the IEEE*, **69**, 529–541.

Portilla, J. and Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International journal of computer vision*, **40**, 49–70.

Rakvongthai, Y. and Oraintara, S. (2008). On phase statistics of complex wavelet coefficients at edges. *European Conference on Signal Processing*, pages 2–6.

Ramírez-Cobo, P., Lee, K. S., Molini, A., Porporato, A., Katul, G., and Vidakovic, B. (2011). A wavelet-based spectral method for extracting self-similarity measures in time-varying two-dimensional rainfall maps. *Journal of Time Series Analysis*, **32**(4), 351–363.

Selesnick, I. W., Baraniuk, R. G., and Kingsbury, N. G. (2005). The dual-tree complex wavelet transform. *IEEE Signal Processing Magazine*, **22**(6).

Strang, G. and Nguyen, T. (1996). *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA.

Veitch, D. and Abry, P. (1999). A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory*, **45**, 878–897.

Vidakovic, B. (1999). *Statistical Modeling by Wavelets*. John Wiley & Sons, Inc.

Wang, T. and Karayiannis, N. (1998). Detection of microcalcifications in digital mammograms using wavelets. *IEEE Transactions on medical imaging*, **17**(4), 498–509.

Zhang, X.-P., Desai, M., and Peng, Y. (1999). Orthogonal complex filter banks and wavelets: Some properties and design. *IEEE Transactions on Signal Processing*, **47**(4), 1039–1048.