

Chapter 8

A Statistical Classification Model for SAXS Images

This chapter develops a statistical model to estimate the probability that a sample is normal, benign or malignant breast tissue from the coefficients of the adaptive transformation of a SAXS image. The coefficients are produced when the filters described in Chapter 7 are used in the adaptive image transform and applied to the SAXS image data set.

These coefficients can be considered measures of features that can be input into the statistical model to provide a classification of the most probable tissue type for each SAXS image. The samples that are imaged may be deemed to be ‘macro’ in size yet the features being sought are measured on the nano-scale. Each sample is likely to be a mixture of normal, benign and malignant nano-structures. The critical malignant structures may not be dominant in the sample and for those cases, the adaptive image transform coefficients will be predominately those associated with normal or benign features. Yet it is vital that the features indicative of malignant tissue, however slight, are not ignored when determining the tissue type. This classification model is designed to utilise features in the coefficient data set that are indicative of malignancy, even when those features may not be dominant.

Some samples may produce greater evidence of belonging to a certain group than others and as a result will have a greater probability of belonging to particular tissue type. This probability will from hence forth be termed the ‘posterior probability’ of the group; it is the probability that data from a sample belongs to a particular tissue group as determined by the classification model.

The initial process of model development is quite lengthy as it involves the estimation of this posterior probability for each of the many coefficients across each scale and location for all of the filters in the library that was described in Chapter 7. The reason for proceeding in this exhaustive fashion is so that any coefficient that suggests malignancy is noted.

This chapter is organised into four sections. Section 8.1 specifies the objective of the model to be designed and the modeling challenges that must be met. The statistical methods of Section 8.2 use each transform coefficient (from the very large set of such coefficients) to predict the probability that the tissue sample is labeled either ‘normal’, ‘benign’ or ‘malignant’. In other words, the posterior probability of a sample belonging to each group for each coefficient of the transform is estimated. This is the first step in a model that will assess the joint effects of the coefficients. Section 8.3 details the statistical model that assesses these joint effects by combining these individual probability estimates to yield an overall probability that the SAXS image belongs to each group. A summary of these methods in Section 8.4 concludes the chapter.

The vast amount of coefficients produced by the adaptive image transformation needs to be reduced to a manageable set whilst retaining most of the useful information. This distillation is done by dividing the total data set into four subsets: (i) training, (ii) validation (iii) adjustment, (iv) test and by using a series of statistical models.

Starting with the data in the *training* set, the adaptive image transform based upon the library of filter functions produced a multitude of coefficients for each filter, scale and location. The filter function selection algorithm (Section 7.4) reduces these to the optimal set of adaptive image transform coefficients for each scale-location. Nonetheless, we are still left with a very large number of coefficients.

The remaining coefficients are summarised by a probability density function (at each scale-location) which may be parametric or non-parametric. In general, the coefficients that are associated with scatter from amorphous substances within the tissue are represented by parametric probability density functions whilst those capturing SAXS image features are represented by non-parametric probability density functions. Since scatter from amorphous substances may explain some of the differences amongst tissue types, that information may be useful in classification. The bivariate Gaussian probability density function is trialled first but if this is found

unsuitable, a non-parametric probability density function is used. The non-parametric density estimates are derived using the Mexican hat contourlet which reduces storage requirements. The Mexican hat contourlet transform is applied to a two-dimensional histogram of the real and imaginary components of each adaptive image transform coefficient. The magnitude of the resultant Mexican hat contourlet transform coefficients is then modified before back transforming to give the smooth bivariate non-parametric probability density function estimate of the adaptive image transform coefficients for each group. A very detailed explanation of the Mexican hat contourlet smoother is provided in Sections 8.2.4-8.2.8 in order to make the steps to obtain the non-parametric density estimates very explicit.

These results are now used in the *validation set*. Images in the validation set are still grouped as normal, benign or malignant and the coefficients in this set have the same indices (i.e. filter number, scale, location) as those in the training set. Across images for each coefficient at a time, the coefficients, $z_{l,a,x,y}^j$ are transformed into density estimates $\hat{p}_{g_{N\text{adj}}}(z_{l,a,x,y}^j)$, using the probability density functions derived from the training set. These density estimates are now used in Bayes' rule (equation 8.2) to estimate the posterior probability of group membership for each group at each scale-location of the adaptive image transform coefficients of each image. For a single coefficient at a particular scale-location, these probabilities of each of the j images is encoded by the triple $(\hat{\text{Prob}}_n(z_{l,a,x,y}^j), \hat{\text{Prob}}_b(z_{l,a,x,y}^j), \hat{\text{Prob}}_m(z_{l,a,x,y}^j))$. These probability estimates are obtained for each coefficient (at each filter, scale and location) for each of the j images in the validation set. The adaptive image transform coefficients which are potentially the most useful in discriminating amongst groups can be identified as those which have large probability of belonging to the true group.

For each image in the validation set, a naive estimate of posterior probabilities of group membership is obtained by averaging the probabilities across filters, scales and locations. For the j th image, these naive probability estimates are written as

$$(\hat{\text{Prob}}_{j,\text{norm}}, \hat{\text{Prob}}_{j,\text{ben}}, \hat{\text{Prob}}_{j,\text{mal}}). \quad (8.1)$$

The last operation with the validation set is to make a histogram (across images within each group) to summarise the naive posterior probability estimates (equation 8.1) for each group. The histograms are smoothed using the Walsh wavelet packet transform because the functions used in this transform are appropriate for data which take values on the interval $[0, 1]$.

The *adjustment* set serves the purpose of calibrating the smoothed univariate histograms. The purpose of this is to adjust the histograms derived by the naive posterior probability estimates for each group. The adjustment is performed by using the observed data in the adjustment set to correct predictions made from the univariate histograms in the validation set.

Finally the adaptive image transform coefficients of the test data set are used to predict group membership via the functions derived above. These predictions can then be compared with the known classifications to assess the accuracy of the methodology.

The specific statistical contributions of this chapter include:

- i) A probability density function estimation method for large amounts of image data that is termed conditional-or estimation. The term ‘exclusive-or’ is abbreviated by the acronym XOR in this thesis (Section 8.2.2).
- ii) The Mexican contourlet transform: a multi-scale, multi-dimensional transformation of bi-variate data (Section 8.2.4).
- iii) Developing methods of smoothing two-dimensional histograms using the Mexican hat contourlet transform (Sections 8.2.5 and 8.2.8).
- iv) Creating a multi-scale smoother using the Walsh wavelet packet transform for data defined on the interval (Sections 8.3.3 and 8.3.5).

8.1 Modeling Objectives and Challenges

The classification model is a key component of the SAXS diagnostic system and provides a means to infer tissue state. A wide range of classifiers already exist in the literature including:

- a) discriminant analysis (Fischer 1936; Rao 1948; Lachenbruch & Goldstein 1979; Hand 1982; Friedman 1989; Hastie, Buja & Tibshirani 1995; Biernacki *et al* 2005; Li & Yuan 2005; Zhu & Martinez 2006)
- b) nearest-neighbour classifiers (Fix & Hodges 1951; Cover & Hart 1967; Dasarathy 1991; Hastie & Tibshirani 1996)
- c) support-vector machines/machine learning (Agrawala 1977; Vapnik 1995; Cristianini & Shawe-Taylor 1999; Abe 2005; Rossi & Villa 2006)
- d) neural networks (Rumelhart & McClelland 1986; Hertz, Krogh & Palmer 1991; Cheng & Titterton 1994; Bishop 1995; Haykin 1999)
- e) tree classifiers (Morgan & Sonquist 1963; Morgan & Messenger 1973; Breiman *et al* 1984; Buntine 1992; Breiman 2001)

Many of these classifiers could be applied as ‘off-the-shelf’ technology and used to classify the SAXS images. This has already been done by Butler *et al* (2003), Erickson (2005) and Round (2006), such an approach misses an opportunity to design a classification model suited to the structure of the data at hand. Given the important and central role of the classifier it is prudent to focus upon the development of a model that is not only accurate but also interpretable.

The objective of the statistical model to be developed is to *accurately identify the pathology of the tissue sample* based upon the observed SAXS image. A model that simply assigns a label (for instance: normal, benign or malignant) given image feature data is not sufficient. Estimates of the probability that the sample belongs to each of these three groups is required, medical experts will then be able to use this information in the wider context to determine if there is a need to do further assessment. The assessment of the model is not then as simple as determining the number of samples correctly identified on a future test or validation set (as done in Butler *et al* (2003), Erickson (2005), Round (2006)), but includes a detailed consideration of how well the model is estimating the probabilities of group membership. The problem is at heart one of regression, using the adaptive transform coefficients to predict the probability that the tissue belongs to a particular group. This approach allows incorporation of the concept that some samples may present

greater evidence of belonging to a certain group than others. For instance, two breast tissue samples might be malignant but belong to different histological stages. The more advanced histological stage tissue sample might have more evidence of collagen structure alteration than the less advanced sample. Using a statistical model to estimate the probability of group membership allows for this possibility and others. A ‘hard’ classification that assigns a label (say ‘normal’, ‘benign’ or ‘malignant’) to a tissue sample under investigation might completely fail to detect a situation in which only a small proportion of the tissue sample is malignant, the rest being either benign or normal healthy tissue. In such a situation the majority of the scatter recorded on the SAXS image will be from either normal or benign tissue and the contribution from the malignant tissue might be obscured. Estimation of the probability of group membership allows this evidence (however slight to be recorded) and the experts alerted to the risk associated with the sample. Statistical scientists are very familiar with the concept of estimating probabilities in classification problems (as evidenced by Ripley, 1996), but the concept seems to have been missed in the SAXS image analysis fraternity. The model developed in this thesis advocates the use of the probability of group membership, rather than the ‘hard’ classification to infer the state of the tissue.

The modeling of the probability of membership tissue group given the adaptive image transform coefficients of breast tissue SAXS images presents several challenges. The most demanding of which are:

- i) High dimensionality: a large number of transform coefficients might provide important diagnostic information that needs to be incorporated in the classification model.
- ii) Limited training sample size: a relatively small number of SAXS images from unique tissue samples combined with a large number of input features (coefficients) can lead to inaccurate model estimates. This problem might have manifested itself in the models presented by Butler *et al* (2003) and Erickson (2005).
- iii) A need and desire to capture the main complexities and dependencies between transform coefficients. This desire drives investigations using non- or semi-parametric techniques in which a more general modeling strategy is sought that is applicable for wider SAXS image analysis applications. We seek a strategy that allows not only the statistical properties of the transform coefficients of SAXS images of normal, benign and malignant breast tissue to be modeled, but also the statistical properties of transform coefficients of SAXS images of different breast tissue grades and brain cancer samples to be modeled. In other words,

we seek a unified cohesive framework to model the transform coefficients of SAXS images in general.

- iv) The opportunity to model the SAXS image across scales using the adaptive image transform.
- v) A scientific need to understand how the transform coefficients and hence both the image and physical structure influence the classification. The ability to understand and interpret the classification model is viewed as a very positive attribute in this context.

This thesis will advocate a modeling strategy in which each adaptive image transform coefficient is first considered individually, allowing group conditional probabilities to be assigned to each relevant filter, scale and location of the transform. Such a strategy is adopted so that vital evidence is not lost, even when that evidence is muted. The classification algorithm that is developed combs deep and wide through all filters, scales and locations of the coefficients to provide the best chance of discovering every useful grain of diagnostic information. This approach will also allow users to identify those coefficients most indicative of group membership (tissue state) and to associate those coefficients with particular image and hence collagen structural features.

The group conditional probabilities from each of the adaptive image transform coefficients will then be collated into a naive overall group conditional probability estimate, which will be further adjusted to account for the fact that the group conditional probabilities at nearby scale-locations may not be mutually exclusive. This adjustment is performed by first estimating probability density functions of the naive overall group conditional probabilities for each group. These probability density function estimates are then modified to ensure the maximum number of correct classifications (by assignment of an observation to the group with the highest posterior probability) on an independent adjustment data set. It is these adjusted probability density functions that are used to estimate the overall group posterior probabilities which are used to diagnose the tissue state of SAXS image observations from an independent test data set.

8.2 Statistical Modeling of the Transform Coefficients

This section is divided into ten components.

- (i) Section 8.2.1 describes how the posterior probabilities of each group at each filter, scale and location of the coefficients of the adaptive image transform of each SAXS image can be estimated.
- (ii) Section 8.2.2 presents the XOR estimation model, that allows either parametric or non-parametric probability density functions to be fit for the adaptive image transform coefficients at different filters, scales and locations.
- (iii) Section 8.2.3 justifies the use of the bivariate Gaussian distribution to model the parametric component of the XOR estimation model.
- (iv) Section 8.2.4 introduces a modified version of the contourlet transform called the Mexican hat contourlet transform.
- (v) Section 8.2.5 describes generally how the Mexican hat contourlet transform can be used for non-parametric probability density function estimation.

- (vi) Section 8.2.6 models the raw two-dimensional histograms that are used to obtain the non-parametric probability density functions as a combination of deterministic (smooth) and stochastic (rough) components.
- (vii) Section 8.2.7 details the practical method for estimating probability density functions non-parametrically using the Mexican hat contourlet transform.
- (viii) Section 8.2.8 specifies how to modify the coefficients of the Mexican hat contourlet transform of a two-dimensional histogram in order to obtain a smooth non-parametric probability density function estimate.
- (ix) Section 8.2.9 explains how the individual probability estimates at each filter, scale and location index of the adaptive image transform of a SAXS image are collated into probability arrays.
- (x) Section 8.2.10 provides an overall summary to this section.

The overall achievement of Section 8.2 is a method that allows estimation of the posterior probability for each group at each coefficient filter, scale and location index from the adaptive image transform of a SAXS image.

8.2.1 Probability Estimates Using Coefficient Information Independently

The objective of this section is to discuss how to reliably estimate the posterior probability of each group for each coefficient of the adaptive image transform. The data set of interest consists of a set of *training* image, T_I . This set can be indexed according to group as follows,

Image	Group	Filter	Scale	Location	Coefficient
1	g_1	1	a_1	(x_1, y_1)	z_{1,a_1,x_1,y_1}
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
N_T	g_n	L	a_m	(x_{N_1}, y_{N_2})	$z_{L,a_m,x_{N_1},y_{N_2}}$

For the filter functions studied in this thesis each coefficient, $z_{l,a,x,y}$, of the super-array of coefficients \mathcal{D}^* (refer to page 175, Section 7.4) contains a real, $\Re(z_{l,a,x,y})$ and an imaginary, $\Im(z_{l,a,x,y})$ component. Recall that equation (7.54) which gives the posterior probability for each group given the relevant coefficients (from the set $R = [\mathbf{z}_{l,a,x,y}(r, 1), \dots, \mathbf{z}_{l,a,x,y}(r, N_g)]$ of the adaptive image transform coefficients form the group g , with indices (l, a, x, y)) is,

$$\Pr[G = g_{N\text{adj}} | \mathbf{z}_{l,a,x,y}(r) = z, r \in R] = \frac{\pi_{g_{N\text{adj}}} p_{g_{N\text{adj}}}(z)}{\sum_{l=1}^n \pi_{g_{n^*}} p_{g_{n^*}}(z)} \quad (8.2)$$

(adapted from equation 2.4, page 19, Ripley 1996). Accurate determination of this probability requires accurate estimates of the probability density functions, $p_{g_{n^*}}(z)$, these estimates are denoted as $\hat{p}_{g_{n^*}}(z)$. Similar probability density function estimates were required in Section 7.4 to estimate the probability of misclassification. In this case there are potentially an immense number of probability density functions that need to be estimated, one for each filter, scale and location. The methodology requires that all probability density functions be estimated for calculation of the probability of misclassification in the filter function selection algorithm but the XOR estimation model reduces the number of probability density functions that need to be *stored* in the final model.

Many methods exist for density estimation including:

- (i) histogram estimators (Sturges 1926; Tarter & Kronmal 1976; Scott 1979; Freedman & Diaconis 1981; Chambers *et al* 1983; Fox 1990; Wand 1997)
- (ii) kernel smoothing (Rosenblatt 1956; Whittle 1958; Parzen 1962; Devroye & Györfil 1985; Silverman 1986; Härdle 1991; Sheather & Jones 1991; Terrell & Scott 1992; Wand & Jones 1995)
- (iii) orthogonal series (Schwartz 1967; Watson 1969; Tarter & Kromal 1970; Brunk 1978; Wahba 1981)
- (iv) wavelet nonparametric smoothing (Masry 1994; Hall & Patil 1995a; Donoho *et al* 1996; Penev & Dechevsky 1997; Mueller & Vidakovic 1998).

All of these techniques could be applied to the data set of adaptive image transform coefficients. One challenge that must be addressed is that the posterior probabilities for each group must be estimated and stored for each coefficient in the transform. Because each coefficient consists of both a real and an imaginary component, bivariate probability density functions are required. This can be a computationally demanding problem. For instance, the SAXS images used in this thesis contain 262,144 (512 x 512) pixels. Assuming that all of the coefficients of the adaptive image transform are retained, then at least $g_n \times a_m \times L \times 262144$ (or $g_n \times a_m \times L \times N_1 \times N_2$ for g_n groups, a_m scales, L filters and image dimensions of size N_1 by N_2 in general) bivariate probability density must be estimated and stored prior to the calculation of the posterior probabilities for each group.

8.2.2 The XOR Probability Density Function Estimation Model

This section proposes a practical solution to the estimation and storage of a large number of probability density functions. The solution is called the XOR estimation model and it is used to represent a probability density function either parametrically or non-parametrically. At each scale-location, the transform coefficients from the training data set T_I are summarised by probability density functions. The coefficients are those from the filters chosen by the filter function selection algorithm. Thus at each scale-location, the information from the SAXS image is condensed into one, two or possibly three probability density functions of coefficients from the transform which is designed to extract specific features at that scale-location.

Where the probability density functions can be represented parametrically, there is compression and reduced data storage. However, not all probability density functions are suitable for parametric modeling and non-parametric estimates of the probability density functions are used in these cases. The cost of using the non-parametric probability density function estimates is increased computation time and storage.

The reasons for having a mix of parametric and non-parametric densities are that:

- (i) Most of the adaptive image transform coefficients are associated with amorphous scatter and these are reliably represented by bivariate Gaussian probability density functions. This is a succinct way of summarising those coefficients.
- (ii) Those coefficients that are associated with distinguishing SAXS image features are not usually well represented by bivariate Gaussian probability density functions. Were the desire for minimum storage of the probability density function estimates to take precedence, the utility of that feature in calculating the group membership (of the SAXS image) would be compromised.

The first move is to estimate the parametric probability density functions for each set (over images) of adaptive image transform coefficients. If this probability density function estimate is satisfactory, the parameter estimates of the probability density function are saved and the XOR estimation algorithm jumps to the next coefficient. If the estimated bivariate Gaussian density is unsuitable (as determined by a statistical measure), then a non-parametric probability density function estimate is calculated in a way that avoids excessive storage.

The XOR estimation model represents each probability density function $p_{g_{n^*}}(z)$ from each group g_{n^*} and for each coefficient $z_{l,a,x,y}$ as,

$$\begin{aligned} \hat{p}_{g_{n^*}}(z_{l,a,x,y}) &= \hat{p}_{\text{Parametric}}[\mathfrak{R}(z_{l,a,x,y}), \mathfrak{S}(z_{l,a,x,y})|g_{n^*}] \\ &\underline{\vee} \hat{p}_{\text{Non-parametric}}[\mathfrak{R}(z_{l,a,x,y}), \mathfrak{S}(z_{l,a,x,y})|g_{n^*}] \end{aligned} \quad (8.3)$$

where the $\underline{\vee}$ symbol denotes the logical operation of exclusive disjunction, which means that the probability density function estimate $\hat{p}_l(z_{l,a,x,y})$ can be either $\hat{p}_{\text{Parametric}}(\mathfrak{R}(z_{l,a,x,y}), \mathfrak{S}(z_{l,a,x,y})|g_{n^*})$ or $\hat{p}_{\text{Non-parametric}}(\mathfrak{R}(z_{l,a,x,y}), \mathfrak{S}(z_{l,a,x,y})|g_{n^*})$ but not both.

Parametric estimates of the probability density function are preferred when the number of samples in the training data set is limited. The disadvantage of parametric models (and hence the need for non-parametric models) is that it is possible that they may miss important aspects of the probability density functions and hence obscure key diagnostic aspects of the data.

In this thesis a bivariate Gaussian distribution was used as a model of the parametric version of the probability density function. Selection between the parametric or non-parametric version of the probability density function estimate for a particular group and coefficient is guided by statistical tests to determine how well the data fits a certain parametric model. For instance, the data could be assessed to determine how good a description the bivariate Gaussian distribution fits. Ascertaining the suitability of either parametric or non-parametric probability density function estimates requires a large number of Shapiro-Wilk tests. This can lead to an unacceptable number of false positives for either type of density estimate. To optimise storage and sustain accuracy of prediction, Benjamini & Hochberg's (1995) false discovery rate method was used to control the expected proportion of errors among the rejected hypotheses (parametric or non-parametric) thereby improving the reliability of selecting the appropriate type (parametric or non-parametric) probability density functions.

8.2.3 Parametric Probability Density Functions of the XOR model for SAXS Images.

This section explains that the bivariate Gaussian probability density function is suitable for the parametric component of the XOR estimation model. Figures 8.1(a)-(f) displays the univariate kernel density estimates for real and imaginary components of the coefficients of the inner product of SAXS images using the Chebyshev filter function. The plotted probability density function estimates were obtained using unbiased cross-validation to select the bandwidth of the Gaussian kernel smoother for select coefficients of the transform for SAXS images ($n = 50$) produced by malignant breast tissue. Coefficient locations for this figure were selected to correspond to particular image features; these features were the amorphous scatter, the axial scattering ring and the equatorial scattering ring. The probability density function estimates of the real and imaginary components of the coefficient [scale $a = 1$, location $(x, y) = (2, 2)$] that is associated with the amorphous scatter are displayed in Figures 8.1(a) and (b).

Visual inspection suggests that both components of the coefficients can be modeled using a Gaussian distribution. The Shapiro-Wilk test applied to the real and imaginary components of the coefficients separately can be used to assess the fit of the Gaussian model (Shapiro & Wilk 1965; Royston 1982; Royston 1982; 1992). It compares the null hypothesis H_0 that the Gaussian distribution fits the data to the alternative hypothesis H_a that the Gaussian distribution does not fit the data well. If the null hypothesis is accepted for both components then the bivariate Gaussian distribution and hence a parametric estimator is used, otherwise a non-parametric estimator is used to calculate the probability density function estimate.

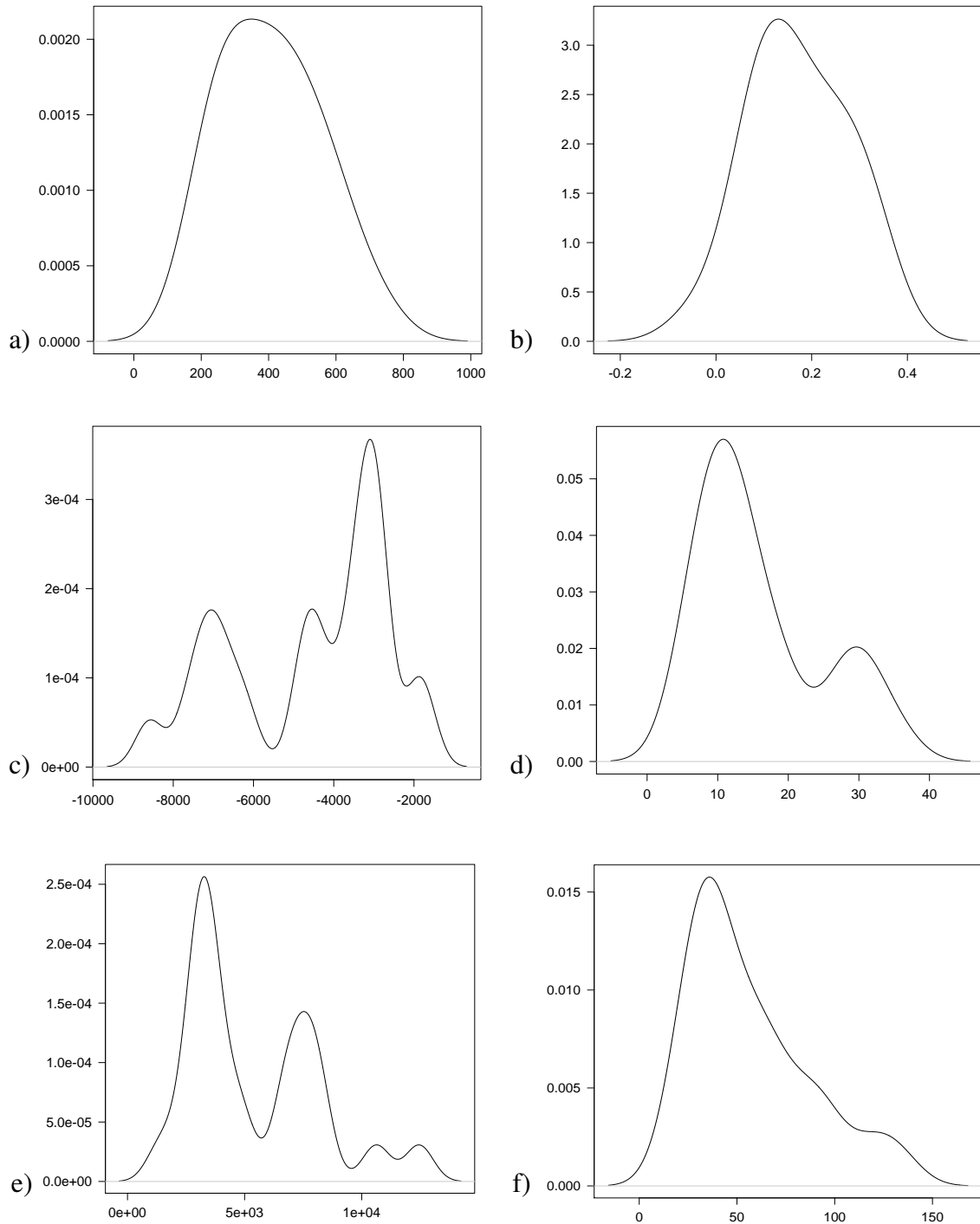


Figure 8.1: Kernel density estimates of select coefficients of the inner product of the Chebyshev filter with the SAXS images of malignant tissue. Figures (a) and (b) correspond to the real and imaginary components of an amorphous scatter feature, figures (c) and (d) to the real and imaginary components of an axial scattering ring feature and similarly (e)-(f) to an equatorial scattering ring feature.

The statistical assessments using Shapiro-Wilk tests across all data support the hypothesis that the bivariate Gaussian distribution models well the distribution of those coefficients associated with amorphous scatter. For instance, the coefficient formed by the Chebyshev filter for parameters [scale $a = 1$, location $(x, y) = (2, 2)$] produced Shapiro-Wilk tests with p-values of 0.44 and 0.59 respectively. In contrast, the density estimates of coefficient magnitude in Figures 8.1(c) and (d) appear to be multi-modal. The coefficients extracted for use in this figure correspond to scale-location [$a = 1$, $(x, y) = (256, 382)$] and are associated with an axial scattering ring feature. In this particular case, Shapiro-Wilk tests report p-values of 0.05 and 0.01 respectively - providing some evidence that a Gaussian distribution is a poor model of this data. The density estimates displayed in Figures 8.1(e) and (f) are of coefficients extracted from scale-location [$a = 1$, $(x, y) = (185, 256)$] which corresponds to the location of an equatorial scattering peak. A Gaussian distribution might also be a poor fit in this case, as the Shapiro-Wilk test produces p-values of 0.01 and 0.04 respectively.

In all cases examined those coefficients that correspond to image locations that describe amorphous scatter have magnitudes that are well modeled by a Gaussian distribution. For those image locations corresponding to scattering (meridional or equatorial) rings, the Gaussian distribution does not appear to fit well. Space does not permit a full presentation of the density estimates of each coefficient (across scales and positions), for each tissue group and each filter function of the transform. Furthermore, in some instances it was difficult to reliably associate coefficient location with image feature. The three locations [$a = 1$, $(x, y) : (2, 2), (256, 382), (185, 256)$] required expert knowledge of the problem to reliably associate them with a particular image feature and the coefficients extracted are still subject to (however unlikely) mislabeling errors. The Shapiro-Wilk tests used to assess the fits of Gaussian distributions to this data must only be considered as a guide to model development. The results of the Shapiro-Wilk tests are only presented to provide a numerical assessment of the plots in Figures 8.1(a)-(f) and should not be interpreted as an all encompassing test that determines whether a Gaussian distribution should be fit to the data.

With these limitations in mind, the following general trend has been observed (when coefficient location and image feature can be associated), the Gaussian distribution does not describe the magnitude of the coefficients well when these coefficients describe scattering ring image features. It appears that *non-gaussianity* in the density of coefficient magnitude is indicative of particular features of interest. This property can be exploited in the XOR estimation model

of coefficient densities. Parametric components are described by bivariate Gaussian distributions and are used to describe amorphous (or non-visual, background) image structure and the non-parametric components are used to describe the scattering ring (or visual, non-stationary, image edge) structures. Both the amorphous and the scattering ring structures are of interest for diagnosis. We are interested in both of these features because:

- (i) Both Lewis *et al* (2000) and Fernández *et al* (2002) reported changes in the axial ring scattering feature for malignant breast tissue that are believed to be associated with the axial D-repeat of collagen.
- (ii) Both Fernández *et al* (2004) and Round (2006) reported increases in the intensity of the amorphous scatter feature in SAXS images of malignant breast tissue.

Therefore both the parametric and the non-parametric probability density functions of the coefficients in the data set are important to consider when creating a diagnostic model.

The XOR estimation model approach to probability density function estimation provides a conceptual framework that might assist the modeler with interpretation of the results, the parametric probability density functions in the model correspond to the amorphous scatter features and the non-parametric probability density functions in the model correspond to the scattering ring features in the SAXS images. The model might also allow highly compressed representations of the probability density functions. Visual inspection of the SAXS images, such as those in Figures 2.3(a)-(c), suggest that a large proportion of the pixels from the SAXS images do not correspond to scattering ring features. Coarse estimates obtained by setting those image intensities less than a threshold (in this case a magnitude of seventy five) and counting the number of non-zero pixels suggests that no more than twenty five percent of the image pixels are associated with scattering ring features.

It is therefore reasonable to expect (in this case) that a large number of the coefficients of the adaptive transform of the SAXS image will be well modeled by bivariate Gaussian distributions. This property allows many coefficients to be accurately described by a four parameter vector $\mathcal{N}_{z_{l,a,x,y}|g_{N^{\text{adj}}}} = [\mu_{\Re(z_{l,a,x,y})}, \mu_{\Im(z_{l,a,x,y})}, \sigma_{\Re(z_{l,a,x,y})}^2, \sigma_{\Im(z_{l,a,x,y})}^2]$, that is a succinct description of the bivariate Gaussian distribution. It is difficult to estimate the number of coefficients that will be described by the Gaussian parameter vector $\mathcal{N}_{z_{l,a,x,y}|g_{N^{\text{adj}}}}$ in general, because this depends on the properties of the SAXS images in the N^{adj} th group as well as the number and type of coefficients retained during the filter function selection process of the transform as described in Section 7.4.

8.2.4 The Mexican Hat Contourlet Transform - A Generic Description

This section develops and discusses the Mexican hat contourlet transform for use in the generic smoothing of bivariate data. This transform will later be used for non-parametric estimation of probability density functions in the XOR estimation model of equation (8.3).

The classification of tissue types using the coefficients from transforms of the image data is done using probability density functions of the coefficients in Bayes' rule. Non-parametric probability density functions are not described as succinctly as parametric probability density functions. With such a large number of coefficients coming from the images, it is essential to find a sparse representation of the non-parametric probability density functions and we achieve this with the Mexican hat contourlet transform. This can substantially reduce the storage requirements of the SAXS image classification model.

The contourlet transform was discussed in Section (5.3.3) where it was stated that it offered a sparse representation of images. Other bivariate digital data can also be represented with relatively few (compared to the number of data entries) contourlet transform coefficients. Therefore the contourlet transform can be used to provide highly compressed representations of bivariate data. Because it provides a sparse representation of bivariate data, it can be inferred that a large proportion of the energy (sums of squares of the coefficients) is contained within a few large magnitude coefficients. This implies that many of the contourlet transform coefficients can be discarded without substantial information about the bivariate data being lost. Furthermore, the large magnitude contourlet coefficients can be understood to be capturing the general trends in the data and the small magnitude ones capturing spurious oscillations or 'noise'. As such the contourlet transform is useful for the smoothing of digital bivariate data.

Despite the sparse representation that the contourlet transform offers it still produces a redundant representation of the data. This is due to the multi-scale step of the transform which is based upon the Laplacian pyramid. This multi-scale step results in there being up to as many as $4/3$ times more coefficients than data entries. This increased redundancy translates into increased computation time and is unnecessary. On first consideration the contourlet transform does not appear to be the best method to use for non-parametric estimation of probability density functions. A modification of the contourlet transform is proposed that overcomes this redundancy problem.

The inherent redundancy of the contourlet transform is easily removed using a two-dimensional wavelet function (denoted $\psi_{j,k_1,k_2}(x, y)$) instead of the Laplacian pyramid in the multi-scale step of the contourlet transform. This two-dimensional wavelet function is a special case of the integral wavelet transform functions $\psi_{a,\mathbf{b},\theta}(x, y)$ (with parameters a denoting scale, \mathbf{b} location and θ rotation/orientation) of Section 5.2.1. The key difference between the two functions is that the parameters of $\psi_{j,k_1,k_2}(x, y)$ are set so that they obey critical sampling (refer to equation 5.15). The scale parameter $a \in \mathbb{R}^+$ of the integral wavelet transform functions is restricted to a subset $a_r \subset a$ such that $a_r = 2^j$ where $j \in \mathbb{N}$. Similarly the location parameters $\mathbf{b} = (b_1, b_2) \in \mathbb{Z}^2$ of the integral wavelet transform functions are equivalent to the parameters $\mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2$ when $b_1 = k_1 2^j$, $b_2 = k_2 2^j$. The rotation parameter θ of the integral wavelet transform functions does not exist for the wavelet function $\psi_{j,k_1,k_2}(x, y)$ because this function is held at a fixed orientation throughout the analysis. This is equivalent to setting $\theta = C$ where $C \in [0, 2\pi)$ for the integral wavelet $\psi_{a,\mathbf{b},\theta}(x, y)$.

One possible choice for the wavelet function $\psi_{j,k_1,k_2}(x, y)$ is the Mexican hat wavelet which was first specified in equation (7.3). The mother wavelet function (from which all other re-scaled and translated versions of the Mexican wavelet are derived) can be described in terms of spatial coordinates as,

$$\psi(x, y) = [2 - (x^2 + y^2)] \exp[-\frac{1}{2}(x^2 + y^2)]. \quad (8.4)$$

The wavelet function $\psi_{j,k_1,k_2}(x, y)$ can then be determined from $\psi(x, y)$ using,

$$\psi_{j,k_1,k_2}(x, y) = \frac{1}{|j|} \psi\left(\frac{x - k_1}{j}, \frac{y - k_2}{j}\right). \quad (8.5)$$

(equation 3.1, page 44, Vidakovic 1999). The term $\frac{1}{|j|}$ is a common variant of the normalisation constant $C_\psi^{-\frac{1}{2}}$ that is used to ensure the energy (sums of squares of coefficients) is preserved across scales (page 55, Addison 2002). The Mexican hat wavelet function allows very precise determination of the location of features in bivariate data at the cost of precision in the estimate of its scale. In other words, coefficients (produced by the inner product of the function and the wavelet) have a very localised response to a feature in space but a less local response in scale. As a result, the large magnitude wavelet coefficients that are associated with a feature will then tend to persist across scales.

The search for a suitable smoother resulted in the choice of the Mexican hat contourlet transform. This result came from comparing the Mexican hat contourlet transform coefficients of data simulated from bivariate distributions with known mean and variance to those coefficients from the Mexican hat contourlet transform of random noise. Random noise seldom transforms into large magnitude coefficients and so the Mexican hat contourlet transform coefficients that are of large magnitude most likely arise from signal.

It is preferable to have a few large magnitude coefficients (that are associated with a particular feature in the data) distributed across several different scale bands rather than many smaller magnitude coefficients within the same scale band. The Mexican hat wavelet function provides this capability and this is the reason why it was selected to modify the multi-scale step of the contourlet transform. A wide range of other wavelet functions could also be substituted for $\psi_{j,k_1,k_2}(x, y)$ in the multi-scale step of the transform and doing so would produce a transform with different feature detection properties.

Decomposing data using the Mexican hat contourlet transform across scales is achieved in a manner similar to the separable two-dimensional discrete wavelet transform (Section 5.2.2).

The method is described as follows:

- i) Starting with the parameter $j = J$, convolve the digital representation of the multi-scale filter, $\psi_{J,k_1,k_2}(x, y) \equiv w_1[2k_1]w_2[2k_2]$ with the digital data $f(x, y)$. The rows of $f(x, y)$ are first convolved by $w_1[2k_1]$ and then down-sampled by removing every second entry along the rows. Denote this result \tilde{C}_{J,k_1,k_2} , to the columns of this matrix we apply the filter $w_2[2k_2]$ and down-sample this result by two along the columns.
- ii) This process produces the approximation coefficient matrix C_{J,k_1,k_2} . Continued iterations of the multi-scale are applied for parameters $j = (1, \dots, J - 1)$.
- iii) The ‘pkva’ directional filters that are often used in the ordinary contourlet transform (Phoong *et al* 1995; Do & Vertelli 2001; Po & Do 2006) are then applied to all of the multi-scale coefficient matrices $\{C_{J,k_1,k_2}, C_{J-1,k_1,k_2}, \dots, C_{1,k_1,k_2}\}$ to yield the detail coefficient matrices $\{D_{J,l_J,k_1,k_2}, D_{J-1,l_{J-1},k_1,k_2}, \dots, D_{2,l_2,k_1,k_2}\}$ and the remaining multi-scale coefficient matrix C_{1,k_1,k_2} is retained. Use of the ‘pkva’ filter in the directional step of the Mexican hat contourlet transform ensures that the chance of artificially inducing correlations amongst coefficient magnitudes is minimised (Phoong *et al* 1995).

iv) The full set of coefficients of the Mexican hat contourlet transform is then:

$\{\mathbf{C}_{1,k_1,k_2}, \mathbf{D}_{J,l_J,k_1,k_2}, \dots, \mathbf{D}_{2,l_2,k_1,k_2}\}$ where parameter J denotes the depth of scale selected such that the dimensions of the bivariate data are expressed as $2^J \times 2^J$, parameter l_j denotes the number of directions at scale J , so that $l_j \in (1, \dots, l_J)$ and $k_1 \in \mathbb{Z}$ and $k_2 \in \mathbb{Z}$ indicate location parameters.

Because of the construction of the Mexican hat contourlet transform, the original data $f(x, y)$ can be recovered from the coefficient matrices $\mathbf{D}_{J,l_J,k_1,k_2}$ in a simple way. First we apply the inverse directional filter bank (as per Do & Vertelli (2003) or any other equivalent method) to the $\mathbf{D}_{J,l_J,k_1,k_2}$ matrices to produce a *single* detail coefficient matrix \mathbf{D}_{1,k_1,k_2} . This coefficient matrix is up-sampled by inserting a zero between every coefficient along both the rows and columns to give matrix $\tilde{\mathbf{D}}_{1,k_1,k_2}$. The inverse filter $\bar{\psi}_{J,k_1,k_2}(x, y) = w_1[-(2k_1)]w_2[-(2k_2)]$ is then applied to give the original data $f(x, y) = \mathbf{C}_{J,k_1,k_2}$. Note that reconstruction of $f(x, y)$ could also have been achieved using the \mathbf{C}_{1,k_1,k_2} coefficient matrix in conjunction with repeated iterations of the inverse filters and up-sampling.

8.2.5 Non-parametric Probability Density Function Estimation Using the Mexican Hat Contourlet Transform

8.2.5.1 Overview of Probability Density Function Estimation Using the Mexican Hat Contourlet Transform.

The objective of this section is to smooth the two-dimensional histogram of the real and imaginary parts of the adaptive image transform coefficients, for those coefficients where the bivariate probability density function representation is rejected. The Mexican hat contourlet transform is used to produce estimates $\tilde{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ of bivariate probability density functions $f[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ based upon the two-dimensional histograms $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ that are produced by observations on the magnitude of the adaptive image transform coefficients $[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ in the sample for each group.

The two-dimensional histogram $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ is produced by assigning the ‘frequencies’ that the coefficient magnitudes $[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ occur in the data sample to the locations $[u = \lfloor \Re(z_{l,a,x,y}) \rfloor, v = \lfloor \Im(z_{l,a,x,y}) \rfloor]$ on a data grid which is denoted \mathbf{F}_{raw} . On its own the two-dimensional histogram, $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ is not very useful as in future cases we will need to obtain estimates of the probability density function at previously unobserved values of $[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$. In order to do this, we smooth the two-dimensional histogram $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ and retain this smoothed version as our estimate $\tilde{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ of the probability density function $f[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$. That is,

$$f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})] = \tilde{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})] + \epsilon[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})] \quad (8.6)$$

so that the observed two-dimensional histogram $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ can be expressed as the sum of a ‘smooth’ estimate $\tilde{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ and a stochastic component $\epsilon[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$. The smooth component $\tilde{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ is retained and the stochastic component $\epsilon[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ is discarded.

To facilitate the development of a statistical model (Section 8.2.6) and a method to remove the stochastic component $\epsilon[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$ (Section 8.2.8) the two-dimensional histogram $f_{\text{raw}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$ is re-scaled so that it has a range in the interval $[0, 1]$. This is achieved by first normalising the two-dimensional histogram $f_{\text{raw}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$ by dividing all entries in the histogram by the magnitude of the entry with the maximum value in the histogram, which is denoted $\max\{f_{\text{raw}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]\}$.

The normalised two-dimensional histogram $f_{\text{raw}}^{\text{N}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$ is then found by,
 $f_{\text{raw}}^{\text{N}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})] = f_{\text{raw}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})] / \max\{f_{\text{raw}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]\}$.
 The magnitudes of entries in $f_{\text{raw}}^{\text{N}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$ are then in the range $[0, 1]$. As a consequence of this normalisation, the model in equation (8.6) becomes

$$f_{\text{raw}}^{\text{N}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})] = \tilde{f}^{\text{N}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})] + \epsilon^{\text{N}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})] \quad (8.7)$$

where $\tilde{f}^{\text{N}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$ and $\epsilon^{\text{N}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$ are modified versions of the smooth $\tilde{f}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$ and the absolute value of the stochastic components $\epsilon[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$ (in equation 8.6) that have been re-scaled so that they have a range of $[0, 1]$. To reduce notational complexity, from here on in it is understood that the notations $f_{\text{raw}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$, $\tilde{f}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$, $\epsilon[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$ assume normalisation so that these functions in fact correspond to $f_{\text{raw}}^{\text{N}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$, $\tilde{f}^{\text{N}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$ and $\epsilon^{\text{N}}[\mathcal{R}(z_{l,a,x,y}), \mathcal{S}(z_{l,a,x,y})]$ respectively.

Several methods can be used to produce the estimate $\tilde{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ including kernel density methods (Wand & Jones 1995), but most of these methods will ultimately lead to results that require large amounts of computer memory to store. The Mexican hat contourlet transform offers an alternative solution to this problem, analysing $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ over scales, directions and locations. It achieves this using scaling functions $\phi_{1,k_1,k_2}(x,y)$ that are of fixed size (scale) but that are translated over a variety of positions in the grid as specified by parameters (k_1, k_2) . The Mexican hat contourlet transform also uses detail functions $\rho_{j,l_j,k_1,k_2}(x,y)$ of varying size (as specified by the scale parameter, $j = 2, \dots, J$), at different orientations (as specified by the direction parameter $l_j = 1, \dots, 2^j$) and at different locations on the grid (as specified by location parameters (k_1, k_2)). The analysis of the two-dimensional histogram $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ with the scaling function $\phi_{1,k_1,k_2}(x,y)$ gives coefficients (denoted c_{1,k_1,k_2}) which are stored as entries in the \mathbf{C}_J coefficient matrix. These coefficients are formed by taking the inner product of the two-dimensional histogram, $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ with the scaling function $\phi_{1,k_1,k_2}(x,y)$. That is $c_{1,k_1,k_2} = \langle f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})], \phi_{1,k_1,k_2}(x,y) \rangle$. Each entry of the \mathbf{C}_J matrix corresponds to the c_{1,k_1,k_2} coefficient for a particular grid location that is specified by $(k_1, k_2) = (u, v)$. Analysis of the two-dimensional histogram $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ with each of the detail functions $\rho_{j,l_j,k_1,k_2}(x,y)$ also produces coefficients $d_{j,l_j,k_1,k_2} = \langle f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})], \rho_{j,l_j,k_1,k_2}(x,y) \rangle$ that are stored as entries in the coefficient matrix \mathbf{D}_{j,l_j} . Therefore, the two-dimensional histogram can be represented in terms of the Mexican hat contourlet transform scaling and detail functions as,

$$\begin{aligned} f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})] &= \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} c_{1,k_1,k_2} \phi_{1,k_1,k_2}(x,y) \\ &+ \sum_{j=2}^J \sum_{l_j=1}^{2^j} \sum_{k_1=1}^{2^{-j}k_1} \sum_{k_2=1}^{2^{-j}k_2} d_{j,l_j,k_1,k_2} \rho_{j,l_j,k_1,k_2}(x,y). \end{aligned} \quad (8.8)$$

Given the coefficients $\{\mathbf{C}_1, \mathbf{D}_{2,l_2}, \dots, \mathbf{D}_{J,l_J}\}$ the two-dimensional histogram $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ can be recovered by applying the inverse Mexican hat contourlet transform.

Information in the two-dimensional histogram are captured by the scaling and detail functions, $\phi_{1,k_1,k_2}(x,y)$ and $\rho_{j,l_j,k_1,k_2}(x,y)$ in a different manner. Both are designed to average the frequencies in the two-dimensional histogram over a particular region but differ according to the geometry of the region that is used. The scaling function $\phi_{1,k_1,k_2}(x,y)$ that is used in the first scale of the Mexican hat contourlet transform is actually a Mexican hat wavelet function at scale $a = 1$. Figure 8.2 displays how the $\phi_{1,k_1,k_2}(x,y)$ function extracts data from the two-dimensional

histogram in order to produce a smooth estimate of the probability density function. The scaling function $\phi_{1,k_1,k_2}(x, y)$ centered on location (k_1, k_2) extracts all those entries within a fixed radius r_1 from this centre. The values of the scaling function $\phi_{1,k_1,k_2}(x, y)$ outside this circle are essentially zero. The coefficient c_{1,k_1,k_2} magnitude (which is found by taking the inner product of the Mexican hat wavelet function and the two-dimensional histogram) will therefore be influenced by only those entries in the two-dimensional histogram that fall within this circle. All entries in the two-dimensional histogram that fall outside this circle will eventually be captured by other scaling functions $\phi_{1,k_1^*,k_2^*}, \dots, \phi_{1,k_1^{**},k_2^{**}}$ centered at different locations $[(k_1^*, k_2^*), \dots, (k_1^{**}, k_2^{**})]$ within the grid of the two-dimensional histogram.

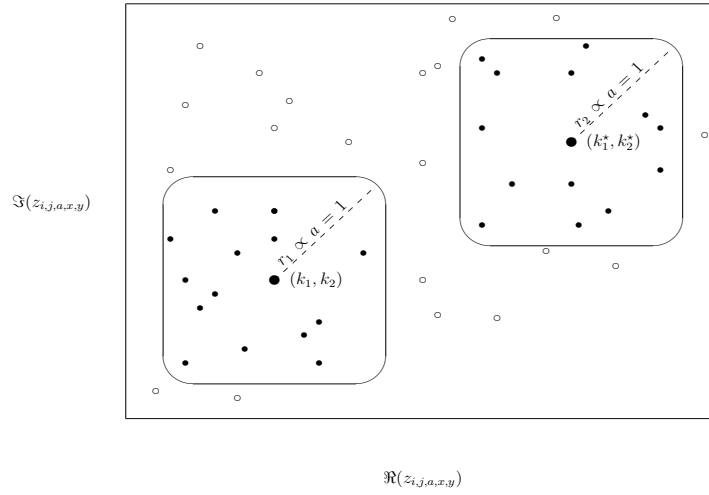


Figure 8.2: The scaling function $\phi_{1,k_1,k_2}(x, y)$ extracts those entries from the two-dimensional histogram $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ that are within a circle of fixed radius, r_1 centered at location (k_1, k_2) . These entries are denoted '•' and it is this information that is used to produce the magnitude of the coefficient c_{1,k_1,k_2} . The function $\phi_{1,k_1^*,k_2^*}(x, y)$ that is of radius r_1 and centered on (k_1^*, k_2^*) also extracts different entries which are used to form the coefficient c_{1,k_1^*,k_2^*} . Other entries in the grid which are denoted '○' are not captured by either of these functions (neither $\phi_{1,k_1,k_2}(x, y)$ nor $\phi_{1,k_1^*,k_2^*}(x, y)$) but will be captured by other scaling functions and stored as coefficients in the coefficient matrix \mathbf{C}_J .

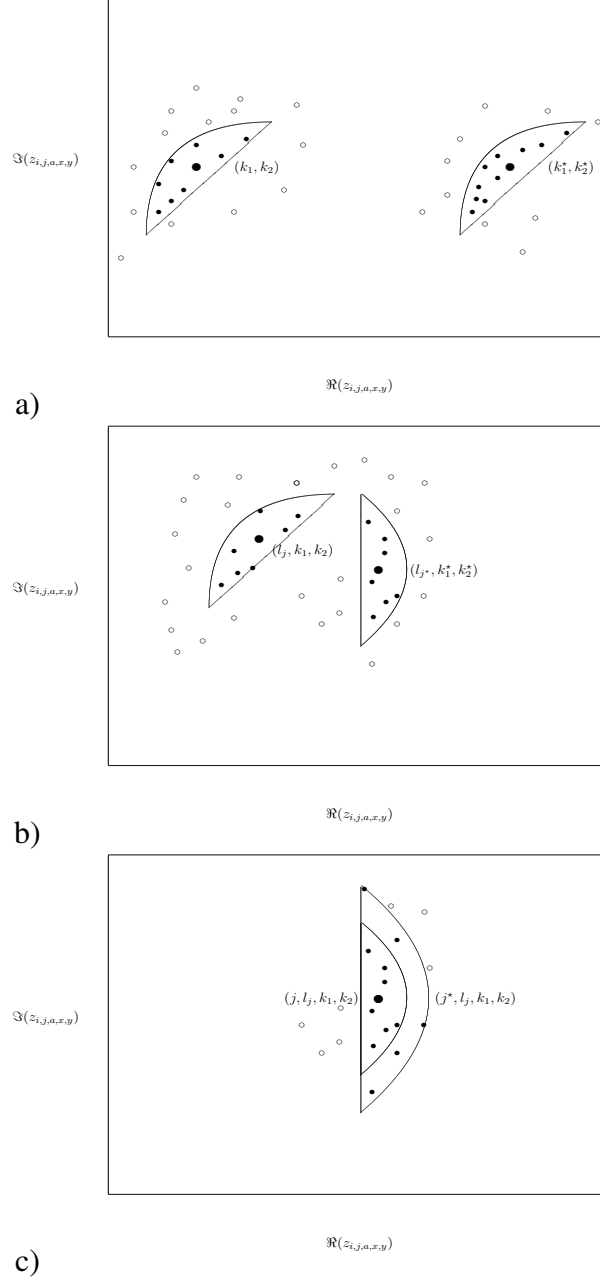


Figure 8.3: Changes in the parameters in the detail functions $\rho_{j,l_j,k_1,k_2}(x,y)$ may result in different entries in the two-dimensional histogram being captured, as denoted by the symbol \bullet . The first figure, Figure (a), demonstrates that changes in the location parameter (k_1, k_2) alter the position of the detail function, Figure (b) demonstrates that changes in the direction parameter $l_j \in 1, \dots, 2^j$ influence the orientation of the detail function and Figure (c) demonstrates that changes in the scale parameter, $j \in 2, \dots, J$ influence the width and length of the detail functions. All of these changes in parameters alter those entries in the two-dimensional histogram that are included in the determination of the coefficient d_{j,l_j,k_1,k_2} .

In contrast, the detail functions $\rho_{j,l_j,k_1,k_2}(x, y)$ extract the information in the two-dimensional histogram $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ using a very different geometry. The non-zero regions of the detail functions $\rho_{j,l_j,k_1,k_2}(x, y)$ resemble arcs rather than circles. Figures 8.3(a)-(c) displays the analysis of the two-dimensional histogram using different detail functions. Thus the coefficient matrices $\{\mathbf{C}_1, \mathbf{D}_{2,l_4}, \dots, \mathbf{D}_{J,l_J}\}$ are considered to encode the results of smoothing the two-dimensional histogram $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ using a different type of smoothing function in each case.

The Mexican hat contourlet transform is designed to provide a succinct representation of bivariate data. It inherits this property from the contourlet transform which has close to optimal representation for bivariate data (page 221, Candés & Donoho 2003; Do & Vetterli 2003). Select Mexican hat contourlet transforms are stored in computer memory as a smooth representation of the two-dimensional histogram. Reconstructing the two-dimensional histogram using only the n -largest absolute magnitude coefficients of the Mexican hat contourlet transform produces an n -term smooth of the two-dimensional histogram which is denoted $f_{\text{raw},n}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$. Candés & Donoho (2003) demonstrated that the squared error of $\|f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})] - f_{\text{raw},n}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]\|$ of the curvelet transform (and hence its discrete counterpart the contourlet transform) will converge as $O(n^{-2})$ as $n \rightarrow \infty$. This rate of convergence is an order of magnitude quicker than that which is possible with the wavelet transform (Candés & Donoho 2003). Therefore in terms of the squared error of an n -term coefficient reconstruction, the contourlet family of transforms (which includes the Mexican hat contourlet transform) is an optimal representation of bivariate data. In other words, the majority of information in the two-dimensional histogram $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ is compressed into only a few relatively large magnitude coefficients. Furthermore, it will be demonstrated in Section 8.2.8.3 that the Mexican hat contourlet transform for a range of probability density functions tends to produce larger absolute magnitude coefficients in the lower scale bands. This means that we can remove many of the lower absolute magnitude coefficients but still retain a good representation of the original two-dimensional histogram of the real and imaginary components of the coefficient $z_{l,a,x,y}$ extracted from the SAXS image using the adaptive image transform. If we order the coefficients by their absolute magnitude and then remove (by setting to zero) the n smallest coefficients, then we will tend to retain those coefficients in the lower scale bands. These lower scale bands correspond to lower spatial frequency information and upon applying the inverse Mexican hat contourlet transform, the probability density function estimate $\tilde{f}_n[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ based upon n non-zero Mexican hat contourlet transform coefficients will be a smoother version of the original two-dimensional histogram $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$.

If instead $N^{\text{adj}} > n$ coefficients in the transform were set to zero, then upon back transformation the estimate $\tilde{f}_{N^{\text{adj}}} [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y})]$ based upon N^{adj} non-zero coefficients will be even smoother than the previous estimate $\tilde{f}_n [\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$. Setting to zero greater and greater numbers of coefficients in the coefficient matrices of the Mexican hat contourlet transform also has the advantage of reducing the number of non-zero coefficients that need to be stored in computer memory. This representation of $\tilde{f}_n [\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ is different from that of $f_{\text{raw}} [\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ (in equation 8.8) because the coefficients have been modified. That is,

$$\begin{aligned} \tilde{f}_n [\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})] &= \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} c_{1,k_1,k_2} \phi_{1,k_1,k_2}(x, y) \\ &+ \sum_{j=2}^J \sum_{l_j=1}^{2^j} \sum_{k_1=1}^{2^{-j}k_1} \sum_{k_2=1}^{2^{-j}k_2} \tilde{d}_{j,l_j,k_1,k_2} \rho_{j,l_j,k_1,k_2}(x, y) \end{aligned} \quad (8.9)$$

where $|\tilde{d}_{j,l_j,k_1,k_2}| \leq |d_{j,l_j,k_1,k_2}|$. The method for modifying the coefficients d_{j,l_j,k_1,k_2} to produce the smoothed coefficients $\tilde{d}_{j,l_j,k_1,k_2}$ will be the subject of Section 8.2.8.

A problem when using the Mexican hat contourlet transform for probability density function estimation is the presence of negative values in the scaling function $\phi_{1,k_1,k_2}(x, y)$. This can result in negative values in the probability density function estimate $\tilde{f}_n [\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$, which is undesirable as it can result in the assignment of negative probabilities to future observations. The scaling function takes on negative values over certain ranges which are called the negative ‘lobes’ of the function. Two possible approaches to the solution of this problem include:

- (i) Using a filter function that does not take on negative values.
- (ii) Adjusting the probability density function estimate $\tilde{f}_n [\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ to compensate for the negative values in the multi-scale filter of the transform.

The first solution in point (i) cannot be achieved using any wavelet function $\psi_{a,k_1,k_2}(x, y)$, as by definition the wavelet function must integrate to zero. That is $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_{a,k_1,k_2}(x, y) dx dy = 0$ (adapted from page 44, Vidakovic (1999)). This implies that the wavelet function $\psi_{a,k_1,k_2}(x, y)$ must contain an equal amount of positive and negative parts and therefore the first approach above is not a valid solution to the problem. Using wavelet functions $\psi_{a,k_1,k_2}(x, y)$ in the multi-scale step of any modified version of the Mexican hat contourlet transform will not ensure that the probability density function estimate $\tilde{f}_n [\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ does not contain negative values. Therefore, we adopt the second approach to solve the problem and adjust the probability

density function estimate $\tilde{f}_n [\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$. Practical implementation of the probability density function estimation method using the Mexican hat contourlet transform and the method to adjust the estimate to exclude negative values is described in Section 8.2.7. Prior to the specification of this algorithm a model needs to be developed to fully specify the stochastic component, $\epsilon[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$.

8.2.6 A Model for Smoothing the Raw Two-dimensional Histogram

$$f_{\text{raw}} [\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$$

The Mexican hat contourlet transform can be used for the estimation of bivariate probability density functions. The basic principle behind this density function estimation technique is that the analysis functions of the Mexican hat contourlet transform can describe the density functions using only a relatively few coefficients. These coefficients are substantially larger than the other coefficients which do not contain significant information about the probability density function. Equation (8.6) models the two-dimensional histogram data $f_{\text{raw}} [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ in terms of smooth $\tilde{f} [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ and stochastic $\epsilon[\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ components.

We wish to retain the smooth component $\tilde{f} [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ as our estimate of the probability density function $f [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ and remove the stochastic component $\epsilon[\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$. This is achieved by taking the Mexican hat contourlet transform of the normalised two-dimensional histogram $f_{\text{raw}} [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ and modifying the resulting coefficients d_{j,l_j,k_1,k_2} to produce the smoothed coefficients $\tilde{d}_{j,l_j,k_1,k_2}$.

The coefficients are modified in a process called *thresholding*. In this process, coefficients below a fixed value, $\lambda \in \mathbb{C}$, are attenuated or set to zero according to a pre-specified *thresholding policy*. This policy is set by the modeler and is often determined by the statistical properties of the data to be analysed.

The method that is proposed in this thesis models the stochastic component, $\epsilon[\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ of the two-dimensional histogram $f_{\text{raw}} [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ as samples from many different Gaussian distributions. The specific model considers a discrete digital representation of the model that was presented in equation (8.6).

$$\mathbf{F}_{\text{raw}} = \hat{\mathbf{F}} + \mathbf{E} \quad (8.10)$$

where the grid \mathbf{F}_{raw} corresponds to a large number of samples from the two-dimensional histogram $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$, the grid $\hat{\mathbf{F}}$ corresponds to samples from the probability density function estimate, $\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ and the grid \mathbf{E} corresponds to the stochastic component $\epsilon[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ of the model specified in equation (8.6). The data in the grid \mathbf{E} are modeled such that the samples at each location $\mathbf{E}(u, v)$ are independently sampled from a $\mathcal{N}(0, \sigma_r^2)$ distribution. The variance parameter σ_r^2 is random on the interval $[0, 1]$ and is drawn from a continuous uniform $U(0, 1)$ distribution. The continuous uniform distribution $U(0, 1)$ was selected such that the variance of the stochastic component $\mathbf{E}(u, v)$ at each location (u, v) is of the same order of magnitude as the two-dimensional histogram grid $\mathbf{F}_{\text{raw}}(u, v)$ data.

To ensure that we obtain values of $\epsilon[\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ that are consistent with the range of the normalised version of $f_{\text{raw}}[\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$, we must re-scale the \mathbf{E} grid values. After samples have been taken from the $\mathcal{N}(0, \sigma_r^2)$ distribution for every location (u, v) in the \mathbf{E} grid, the absolute value of the magnitudes of the entries in this grid are found and the maximum value $\max[\mathbf{E}(u^*, v^*)]$ at location (u^*, v^*) is recorded. The magnitude of each entry in the $\mathbf{E}(u, v)$ grid is then divided by this maximum value $\max[\mathbf{E}(u^*, v^*)]$ in order to re-scale the entries in the \mathbf{E} grid to the range $[0, 1]$.

The model that has been proposed in equation (8.10) in combination with sampling the magnitude of each entry from a $\mathcal{N}(0, \sigma_r^2)$ distribution states that each pair of entries $\mathbf{E}(u_1, v_1) = \mathbf{E}(u_2, v_2)$ $u_1 \neq v_1, u_2 \neq v_2$ might well be sourced from different Gaussian distributions. We call such a statistical process a heteroskedastic noise model. The model states that for each sample $\mathbf{F}_{\text{raw}}(u, v)$ in the two-dimensional histogram there is a certain amount of uncertainty but the extent of that uncertainty is unknown. All we can say is that the magnitude of each entry $\mathbf{E}(u, v)$ is drawn from a random distribution $\mathcal{N}(0, \sigma_r^2)$. The magnitude (hence the extent) of the variance parameter σ_r^2 for each pair (u, v) in the \mathbf{E} grid is completely random.

The process that generates each of the σ_r^2 parameters involves sampling a continuous uniform $U(0, 1)$ distribution. Use of the continuous uniform $U(0, 1)$ distribution communicates that we are in total ignorance of the uncertainty that should be assigned to each location (u, v) on the grid \mathbf{E} . If we better understood how the uncertainty varied with the parameters (u, v) then a more precise model could be developed to describe the stochastic component more accurately. This is not the case at the present and we need to implement a general model for the values of the entries $\mathbf{E}(u, v)$, such as the heteroskedastic noise model that was introduced in this section. Nonetheless, the heteroskedastic noise model captures key characteristics of the type of stochastic processes that are of interest.

These include:

- (i) A process \mathbf{E} in which samples at different locations (u_1, v_1) and (u_2, v_2) are very likely drawn from different distributions $\mathcal{N}(0, \sigma_{r,1}^2)$ and $\mathcal{N}(0, \sigma_{r,2}^2)$. This can be interpreted as meaning that the uncertainty varies across each location in the \mathbf{E} grid.
- (ii) Generating multiple \mathbf{E} grids - say \mathbf{E}_1 and \mathbf{E}_2 by sampling from $\mathcal{N}(0, \sigma_r^2)$, $\sigma_r^2 \sim U(0, 1)$ will in most cases generate two grids that have different magnitudes at each location (u, v) . In other words, the heteroskedastic model can generate different realisations of the stochastic process/component.
- (iii) It accounts for the randomness in the sampling process in which tissue samples were obtained in order to produce SAXS images. To understand this problem, consider two data sets $T_1 = [I_{1,1}(x, y), \dots, I_{n_1,1}(x, y)]$ and $T_2 = [I_{1,2}(x, y), \dots, I_{n_2,2}(x, y)]$ of SAXS images produced by the tissue samples all of which were (supposedly) drawn independently of each other. Upon taking the adaptive image transform of the images and implementing the methodology that is described in this chapter, we arrive at the probability density function estimates $\tilde{f}_{T_1} [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ and $\tilde{f}_{T_2} [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ formed by the data from training sets T_1 and T_2 respectively. In general, $\tilde{f}_{T_1} [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)] \neq \tilde{f}_{T_2} [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$, because the probability density function estimate depends on the data in the training sets T_1 and T_2 .

Point (iii) above is of great practical importance as the probability density function estimates are used to estimate the probability that an observation belongs to each tissue group. The key issue is that the probability density function estimate $\tilde{f} [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ and hence the probability estimate of the sample belonging each tissue group is dependent on the data sample. This data sample is obtained by sampling the population of subjects within a particular tissue group. For instance, the samples of malignant tissue used in this thesis were obtained by sampling the population of all patients with malignant breast tissue.

Now this sampling process is itself influenced by a whole range of factors such as:

- (i) the hospital(s) from which the samples were obtained.
- (ii) the age of the subjects.
- (iii) the stage (maturity of the cancer) at which the samples were collected.

These factors might in turn be influenced by other epidemiological variables that:

- (i) might bias the distribution of ages from which samples are collected.
- (ii) influence the types of tissue pathologies that occur.
- (iii) determine the hospital(s) participating in the sample collection process.
- (iv) might favour particular stages of cancer being detected.

Such processes are not very well understood and it is very difficult to identify let alone explicitly model the influence of such factors on the sampling process. The solution proposed is to incorporate a certain amount of uncertainty (as specified by the $\mathcal{N}(0, \sigma_r^2)$ distributions) when producing the estimate $\tilde{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ of the bivariate probability density function.

8.2.6.1 Coefficient Thresholding Methodology

A statistical model was required to describe the stochastic component $\epsilon[\Re z_{l,a,x,y}^j, \Im z_{l,a,x,y}^j]$ of the model in equation (8.6). This component was assumed to have been generated by repeated sampling of random $\mathcal{N}(0, \sigma_r^2)$ distributions. Now that the model for the stochastic component has been specified, the Mexican hat contourlet transform can be used to remove its influence and produce the estimate $\tilde{f}[\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$. This is achieved by applying the Mexican hat contourlet transform to \mathbf{F}_{raw} , applying a thresholding policy and then applying the inverse Mexican hat contourlet transform to obtain $\hat{\mathbf{F}}$. This thresholding policy needs to be developed before $\hat{\mathbf{F}}$ can be estimated.

Probability density function estimation methods that combine multi-scale transforms with thresholding policies are not without precedent. Wavelet transform methods have been used by Donoho (1993), Donoho *et al* (1996), Hall & Patil (1995a; 1995b; 1996) and Walter (1994) among others to produce probability density function estimates. These models take advantage of the energy compaction property of the wavelet transform in order to produce succinct representations of the probability density function estimates. The term ‘energy compaction’ refers to the property of both wavelets and contourlets in which the transform of the data contains relatively few large magnitude coefficients. By keeping only the largest magnitude coefficients, wavelet methods have been able to produce succinct (in scale-location space) and accurate probability density function estimates. The Mexican hat contourlet transform offers potentially superior performance over the wavelet methods of bivariate probability density function estimation because it combines the advantages of the wavelet transform with the advantages of the contourlet transform. Incorporating sub-sampling and the Mexican hat wavelet function removes the inherent redundancy in the multi-scale step of the contourlet transform (refer to Section 5.3.3). Whilst retaining the directional component of the contourlet transform (in the Mexican hat contourlet transform) ensures rapid ($O(n^{-2})$ for n coefficients) convergence of the probability density function estimate formed by retaining the n largest coefficients $\tilde{f}_n[\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$ to the ideal representation of the probability density function estimate $\tilde{f}[\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$. This is an order of magnitude faster than the rates possible with the wavelet transform (refer to Sections 5.3.2 and 5.3.3, as well as page 221, Candés & Donoho 2003). Therefore it is likely that methods based upon the Mexican hat contourlet transform can provide probability density function estimates that are more succinct than those methods that are based upon the wavelet transform.

Eslami & Radha (2003) and Varshney (2004) developed thresholding methodology for contourlet transform coefficients. Their methods assumed that the stochastic component of the data (the ‘noise’) could be modeled by a $\mathcal{N}(0, \sigma^2)$ distribution. In the context of probability density function estimation, such models assume that every entry in the \mathbf{E} grid of equation (8.10) is sampled from the *same* $\mathcal{N}(0, \sigma^2)$ distribution. This is very different model from the heteroskedastic noise model that was proposed in Section 8.2.6. The heteroskedastic noise model uses (in general) a different $\mathcal{N}(0, \sigma_r^2)$ distribution to determine the magnitude of each entry in the \mathbf{E} grid. Therefore, the methods of Eslami & Radha (2003) as well as Varshney (2004) are not readily extended in order to develop a thresholding policy of Mexican hat contourlet transform coefficients. Nonetheless, a thresholding policy can be developed based upon the results of simulation studies. These studies generated random matrices of data by sampling $\mathcal{N}(0, \sigma_r^2)$, $\sigma_r^2 \sim U(0, 1)$ distributions and assigning the sampled value to the relevant entry for each matrix. The Mexican hat contourlet transform was applied to these random matrices and parametric (generalised extreme value) distributions were fit to the maximum absolute values of coefficient magnitude. The specific motivations for such a model and the methodology of the thresholding policy will be the subject of Section 8.2.8 of this chapter. Before examining the thresholding policy in detail, we will examine the specific algorithm that can be used to perform non-parametric probability density function estimation using the Mexican hat contourlet transform in Section 8.2.7.

8.2.7 Practical Implementation of the Mexican Hat Contourlet Transform Method for Non-parametric Probability Density Function Estimation

In order to perform non-parametric probability density function estimation using the Mexican hat contourlet transform a practical algorithm is required. The data which are now ingredients to the non-parametric probability density function estimates are those complex coefficients which are stored in the reduced, but still large array \mathcal{D}^* (refer to Section 7.4 on page 178). To develop the classifying model for groups, \mathcal{D}^* is divided into g blocks corresponding to the tissue types from which the images were made (it is three groups in the data that are analysed in this thesis). The diagram of \mathcal{D}^* is reproduced in Figure 8.4, it is annotated to show the division into g blocks in order to aid explanation.

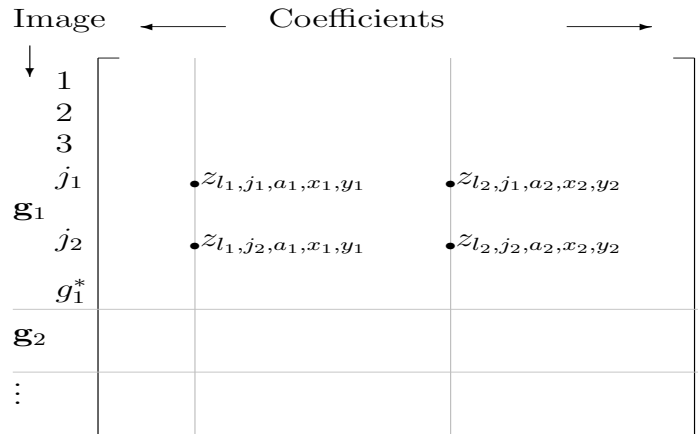


Figure 8.4: Schematic diagram to represent the array \mathcal{D}^* . The coefficients along a column are extracted and sub-divided according to the group to which they belong, these coefficients then serve as the basic data that is input into the non-parametric bivariate probability density function algorithm that is described in this section.

For each tissue type/group, we require adequate sample sizes say $N_g \geq 50$ images from which we will use the corresponding coefficients in \mathcal{D}^* . The strategy is now to calculate a bivariate probability density function estimate over the samples in a group for each filter, scale and location. That is, one at a time, calculate the density of each column of \mathcal{D}^* in Figure 8.4. This estimate will be denoted $\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$.

- (i) When we extract the target coefficients from \mathcal{D}^* , they are collated into a matrix $\mathbf{Z}_{g_{N\text{adj}}}$ which has $N_{g_{N\text{adj}}}$ rows (i.e. the number of samples in the group) and two columns for the real and imaginary parts. For instance the j th SAXS image belonging to the group $g_{N\text{adj}}$ corresponds to the pair of coefficient components $[\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$. The component entry $\Re(z_{l,a,x,y}^j)$ is assigned to the position $\mathbf{Z}_{g_{N\text{adj}}}(j, 1)$ in the data matrix, whilst the component $\Im(z_{l,a,x,y}^j)$ is assigned to the position $\mathbf{Z}_{g_{N\text{adj}}}(j, 2)$ in the data matrix.

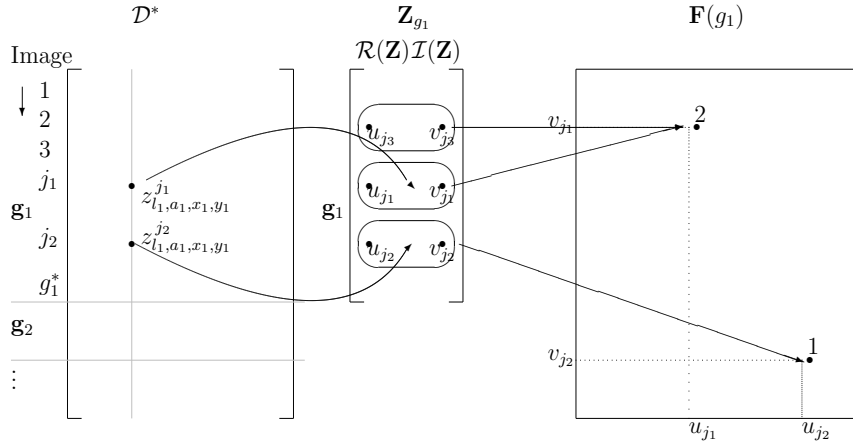


Figure 8.5: Mapping the \mathbf{Z}_{g_1} matrix to the raw \mathbf{F}_{g_1} matrix. Coefficients $z_{l_1,a_1,x_1,y_1}^{j_1}, \dots, z_{l_1,a_1,x_1,y_1}^{g_1^*}$ belonging to the images in the group g_1 from the array \mathcal{D}^* are extracted and placed in the matrix \mathbf{Z}_{g_1} . The rows of this matrix correspond to the j_k th ($k = 1, \dots, g_1^*$) image and the columns correspond to the real $\Re(\mathbf{Z}_{g_1})$ and the imaginary $\Im(\mathbf{Z}_{g_1})$ parts respectively. The entries $u_{j_k} = \Re[z_{l_1,a_1,x_1,y_1}^{j_k}]$ and the imaginary $v_{j_k} = \Im[z_{l_1,a_1,x_1,y_1}^{j_k}]$ of the \mathbf{Z}_{g_1} matrix are then mapped to the raw $\mathbf{F}(g_1)$ matrix with an increment of one being placed at the grid location that is relevant to (u_{j_k}, v_{j_k}) . Coefficients may have the same coordinates (u_{j_k}, v_{j_k}) and the \mathbf{F}_{g_1} matrix will then register as a two at this particular location.

- (ii) In order to get the columns of $\mathbf{Z}_{g_{N\text{adj}}}$ into an amenable format for the Mexican hat contourlet transform it is necessary to map the real and imaginary components of the coefficients onto a two-way grid (magnitude real x magnitude imaginary coefficient) where presence of the sample on the grid is marked by a '1' at the location corresponding to real and imaginary components of the coefficients and a '0' otherwise. This mapping is represented schematically in Figure 8.5 and the grid is referred to as the raw $\mathbf{F}_{g_{N\text{adj}}}$ matrix. The entries in the raw $\mathbf{F}_{g_{N\text{adj}}}$ matrix are all initially zero and are systematically modified with ones for each sample in the $\mathbf{Z}_{g_{N\text{adj}}}$ matrix at the locations indicated by the $\Re(z_{l,a,x,y}^j)$ and $\Im(z_{l,a,x,y}^j)$ components of the coefficients.
- (iii) The dimensions of the raw $\mathbf{F}_{g_{N\text{adj}}}$ matrix are of size 2^{M_1} rows and 2^{M_2} columns. The parameters, M_1 and M_2 are positive integers that are selected to ensure sufficient resolution of the probability density function estimate, $\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$. In this thesis, these parameters and hence the number of data entries in the grid was selected to ensure that the range of possible coefficients values is covered. Denote R_1 as the range of the real components of the coefficients and R_2 as the range of the imaginary components of the coefficients irrespective of group. That is, $R_1 = |\max[\Re(z_{l,a,x,y})] - \min[\Re(z_{l,a,x,y})]|$ and $R_2 = |\max[\Im(z_{l,a,x,y})] - \min[\Im(z_{l,a,x,y})]|$. Similarly calculate the sample standard deviation (denoted s) of each of the components of the coefficients irrespective of group. That is calculate, $s_1 = s[\Re(z_{l,a,x,y})]$ and $s_2 = s[\Im(z_{l,a,x,y})]$, for all of the coefficients (related to the i th filter at the a th scale for the (x, y) -th location) from all of the SAXS images in the data set. Then the parameters are given by $M_1 = \lfloor R_1 + s_1 \rfloor$ and $M_2 = \lfloor R_2 + s_2 \rfloor$, where the notation ' $\lfloor R_1 + s_1 \rfloor$ ' represents the floor function that gives the largest integer less than or equal to $(R_1 + s_1)$. Denote this grid as the 'empty' $\mathbf{F}_{g_{N\text{adj}}}$ matrix. When information related to the magnitude of the real and imaginary components for the coefficients in the data set is added, this matrix will serve as a 'raw' template from which a digital representation of $\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ can be obtained using the Mexican hat contourlet transform.
- (iv) Along each row of the data matrix $\mathbf{Z}_{g_{N\text{adj}}}$ extract the entry belonging to the j th image. Calculate the row index as $I_r = \lfloor \mathbf{Z}_{g_{N\text{adj}}}(j, 1) \rfloor = \lfloor \Re(z_{j,i,a,x,y}) \rfloor$ and the column index as $I_c = \lfloor \mathbf{Z}_{g_{N\text{adj}}}(j, 2) \rfloor = \lfloor \Im(z_{j,i,a,x,y}) \rfloor$. Then in the $\mathbf{F}_{g_{N\text{adj}}}$ matrix at the I_r th row and the I_c th column assign the magnitude one to indicate the location of the data entry. Repeat this process for all $N_{g_{N\text{adj}}}$ entries in the data matrix $\mathbf{Z}_{g_{N\text{adj}}}$.

- (v) The probability density function estimate $\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ is calculated by the weighted grid positions (where the weights are the incidences of coefficients in these positions) according to the Mexican hat contourlet transform. This probability density function estimate $\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ is provided by the $\mathbf{F}_{g_{N\text{adj}}}^g$ matrix which is also called the $\mathbf{F}_{g_{N\text{adj}}}^g$ grid.
- (vi) The probability density function estimate $\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ might take on negative values in those regions of the $\mathbf{F}_{g_{N\text{adj}}}^g$ grid that have few observations. These negative values in the estimate are an artifact produced by the negative values in the ‘lobes’ surrounding the central peak in the Mexican hat wavelet function that is used in the multi-scale step of the Mexican hat contourlet transform. This is undesirable for diagnostic applications because it may produce negative probabilities of a sample belonging to a particular group if it is used in unusually extreme observations. For instance, a sample may produce a SAXS image that when analysed with the adaptive image transform has particular coefficients that would strongly suggest the sample is from malignant tissue. It might be that the data set that is used to estimate the probability density function (for a coefficient at fixed filter, scale and location) of the malignant tissue group has very few observations at these extreme values. To an outside observer it would appear that the sample would be associated with a high probability of being malignant, yet because of the negative values of the probability density function at these extreme magnitudes could produce absurd results, such as a probability estimate of negative value. To overcome this problem, we take the square root of *frequencies* of the $\mathbf{F}_{g_{N\text{adj}}}^g$ grid to get the modified histogram $+\sqrt{\mathbf{F}_{g_{N\text{adj}}}^g}$. We then estimate the transformed density function $\sqrt{\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]}$ using this modified histogram. The inverse transform, $[\sqrt{\mathbf{F}_{g_{N\text{adj}}}^g}]^2$ is then performed to obtain a probability density function estimate $\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ that has only zero or positive values. Other more advanced methods could also be incorporated into this step by modification of the multi-scale filter in the Mexican hat contourlet transform using similar methods to those reported in Hall & Murison (1993), Penev & Dechevsky (1997) or Wand, Marron & Ruppert (1991).

It must also be ensured that the probability density function estimate $\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ integrates to one. That is,

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})] d\Re(z_{l,a,x,y}) d\Im(z_{l,a,x,y}) = 1.$$

Since the Mexican hat contourlet transform does not necessarily obey Parseval's identity (the sum of the squares of the initial data is not the sum of the squares of the coefficient magnitudes), estimating the square root of the probability density function does not guarantee that the probability density function estimate integrates to one as it would with orthogonal wavelet methods (pages 229-230, Vidakovic 1999). A practical solution to this problem was to use the Mexican hat contourlet transform to obtain the estimate $\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ and then to use the relation

$$\int_{-\infty}^{\infty} \int_{-\infty}^{+\infty} \hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})] d\Re(z_{l,a,x,y}) d\Im(z_{l,a,x,y}) \approx \sum_{u=1}^{2^{M_1}} \sum_{v=1}^{2^{M_2}} \tilde{\mathbf{F}}_{g_{N\text{adj}}} (u, v)$$

to set $\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})] d\Re(z_{l,a,x,y}) d\Im(z_{l,a,x,y}) \approx 1$. This was achieved by calculating scaling the $\tilde{\mathbf{F}}_{g_{N\text{adj}}}$ grid by using the entry in this grid with the largest magnitude (denoted $C = \max[\tilde{\mathbf{F}}_{g_{N\text{adj}}}^*(u, v)]$, $u \in 1, \dots, 2^{M_1}$, $v \in 1, \dots, 2^{M_2}$) to produce the re-scaled probability density function estimate $\tilde{\mathbf{F}}_{g_{N\text{adj}}}^*(u, v) = \tilde{\mathbf{F}}_{g_{N\text{adj}}}(u, v)/C$.

- (viii) Apply the Mexican hat contourlet transform to the $\mathbf{F}_{g_{N\text{adj}}}$ matrix using a total of J scales. The lowest scale, $j = 1$ corresponds to the 'smooth' band of coefficients \mathbf{C}_{1,k_1,k_2} , whilst the detail bands at scales $(2, \dots, J)$ are directional filtered versions of the coefficients produced by the analysis of the $\mathbf{F}_{g_{N\text{adj}}}$ matrix with the Mexican hat wavelet.
- (ix) For each coefficient d_{J^*,l_{J^*},k_1,k_2} from the detail coefficient matrix $\mathbf{D}_{J^*,l_{J^*},k_1,k_2}$ for $J^* \in 2, \dots, J$, calculate the relevant scale-dependent threshold $\lambda_{J^*} \in \mathbb{C}$ and apply the relevant thresholding policy (which will be discussed in Section 8.2.8) to produce the attenuated coefficient $\tilde{d}_{J^*,l_{J^*},k_1,k_2}$. Repeating this process for all detail bands $J^* \in (2, \dots, J)$ produces a set of 'smoothed' coefficients $\mathcal{S} = \{\mathbf{C}_{J,k_1,k_2}, \mathbf{D}_{J,l_J,k_1,k_2}, \dots, \mathbf{D}_{2,l_2,k_1,k_2}\}$.
- (x) From the set of smoothed coefficients \mathcal{S} , the inverse Mexican hat contourlet transform is applied by reversing the order of filter operations (as compared to the order in the original decomposition of the Mexican hat contourlet transform) and up-sampling by two at every increase in scale to produce the smoothed digital probability density function estimate $+\sqrt{\tilde{\mathbf{F}}_{g_{N\text{adj}}}}$. This result is then squared to produce the estimate $\tilde{\mathbf{F}}_{g_{N\text{adj}}}$ of the probability density function for the group $g_{N\text{adj}}$ relevant for the i th filter, on the a th scale at location (x, y) .
- (xi) The results are stored and the probability density function estimation process is repeated for all of the g_n groups, the a_n scales, across all locations (x, y) and associated filters from the library \mathcal{L} .

8.2.8 A Thresholding Policy for Mexican Hat Contourlet Transform Coefficients

8.2.8.1 Overview

This section develops the methodology to both estimate a threshold and implement a thresholding policy for coefficients of the Mexican hat contourlet transform. Recall the SAXS image analysis methodology developed to date:

Chapter 6

- (i) A data set of SAXS images of normal, benign and malignant breast tissue samples was obtained.

Chapter 7

- (ii) The adaptive image transform (equation 7.1) of each SAXS image was performed using the filter functions in the library \mathcal{L} (Section 7.3) using the filter function selection algorithm (Section 7.4).

Section 8.2.1

- (iii) Use each adaptive image transform coefficient $z_{l,a,x,y}^j$ (that was retained in the super-array of coefficients \mathcal{D}^* after application of the filter function selection algorithm) independently, the posterior probability of the j th sample in the data set belonging to each group (normal, benign or malignant) was required. This probability was specified in equation (8.2), which required the estimation of $p_{g_{n^*}}(z_{l,a,x,y}^j = z)$ the value of the probability density function of the coefficients $z_{l,a,x,y}^j$ that belong to the group with index g_{n^*} at magnitude $z \in \mathbb{C}$.

Section 8.2.2

- (iv) A probability density function was required to describe the distribution of each coefficient $z_{l,a,x,y}$ in the super-array \mathcal{D}^* for each group in the data set. This requirement imposed that a large number of probability density functions be estimated and stored in computer memory as a component of the breast cancer diagnosis model. The XOR probability density function estimation model was specified in equation (8.3) to reduce the storage requirements for all of the probability density functions.

Section 8.2.3

- (v) The bivariate Gaussian probability density function was advocated where possible, to describe the probability density function of each coefficient $z_{l,a,x,y}$ given group. Where this was not possible, non-parametric probability density functions were required.

Section 8.2.4

- (vi) In order to produce succinct non-parametric estimates of the probability density functions for the coefficients $z_{l,a,x,y}^j$ given group, the Mexican hat contourlet transform was devised.

Section 8.2.5

- (vii) To arrange the data provided by the adaptive image transform coefficients $z_{l,a,x,y}$ into a format amenable to the Mexican hat contourlet transform, two-dimensional histograms $f_{\text{raw}}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ were created. These two-dimensional histograms were on a grid indexed by the real $\Re(z_{l,a,x,y})$ and imaginary $\Im(z_{l,a,x,y})$ components of the coefficients $z_{l,a,x,y}$ respectively. The heights of the two-dimensional histogram corresponded to the frequencies of the the real and imaginary components of the adaptive image transform coefficient magnitudes.
- (viii) The two-dimensional histograms of frequencies of real and imaginary components of adaptive image transform coefficients were modeled in equation (8.6) as being the sum of a smooth $\tilde{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ and stochastic $\epsilon[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ component. The smooth component $\tilde{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ was to be retained as the estimate of the probability density function $\hat{f}[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})] \equiv p_{g_n^*}(z_{l,a,x,y})$ and the stochastic component $\epsilon[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ discarded.

- (ix) Equations (8.8) and (8.9) explained how the two-dimensional histogram $f_{\text{raw}} [\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ could be represented and then modified by taking the Mexican hat contourlet transform and altering the magnitude of the transform coefficients.

Section 8.2.6

- (x) A discrete digital representation of the two-dimensional histogram was presented in equation (8.6). This representation modelled each entry in the \mathbf{E} matrix (the discrete digital representation of $\epsilon[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$) as samples drawn from a stochastic process $\mathcal{N}(0, \sigma_r^2)$, $\sigma_r^2 \sim U(0, 1)$ in which the variance parameter σ_r^2 changes randomly with each draw.

Section 8.2.7

- (xi) Described how to implement the Mexican hat contourlet transform smoother in practice. This smoother requires a thresholding (or weighting) policy as well as a threshold (or weight) in order to modify the Mexican hat contourlet transform coefficient magnitude.

The objective of this section is to develop the methodology to both estimate a threshold and to implement a thresholding policy on the coefficients of the Mexican hat transform of the discrete two-dimensional histogram grid $\mathbf{F}_{N^{\text{adj}}}^g$. Once these objectives have been achieved, it will be possible to perform probability density function estimation using the Mexican hat contourlet transform.

The next two subsections, Section 8.2.8.2-8.2.8.3, demonstrate through simulation how the Mexican hat contourlet can be used to derive a probability density function with a succinct set of coefficients.

Section 8.2.8.2 simulates noise to measure the magnitudes of Mexican hat contourlet transform coefficients for pure noise whereas Section 8.2.8.3 calculates the magnitudes of Mexican hat contourlet transform coefficients from simulated bivariate probability density functions. Knowledge of the coefficient magnitudes of coefficients from: (a) bivariate probability density functions and (b) noise provides a thresholding scheme that will select informative coefficients to make the probability density function estimate.

These experiments compare the coefficients of simulated random matrices (where the magnitude of each entry in the random matrix was drawn from a $\mathcal{N}(0, \sigma_r^2)$, $\sigma_r^2 \sim U(0, 1)$ distribution) to the coefficients of the Mexican hat contourlet transform of bivariate probability density functions. In other words, artificial bivariate data is produced to compare the magnitude of the Mexican hat contourlet transform coefficients of smooth components $\hat{\mathbf{F}}$ to stochastic components \mathbf{E} (refer to equation 8.10). The simulated random matrices are artificial data that represent specific examples of the stochastic component grid \mathbf{E} and the bivariate probability density functions are specific examples of the smooth component grid $\hat{\mathbf{F}}$. The results of these simulation studies indicate that distinct differences exist in the maximum absolute magnitude $|d_{j,l_j,k_1,k_2}|$ for Mexican hat contourlet transform coefficients that are associated with both types of data.

Section 8.2.8.4 exploits the observed differences between Mexican hat contourlet transform coefficients produced by random data and those produced by bivariate probability density functions to develop a coefficient thresholding policy.

The thresholding policy uses a generalised extreme value distribution to model the absolute magnitude of the Mexican hat contourlet transform coefficients (Von Mises 1954). The parameters of the generalised extreme value distribution $p(z, \mu, \sigma, \xi)$ are estimated using the absolute values $|d_{j,l_j,k_1,k_2}|$ of the Mexican hat contourlet transform coefficients of the simulated random matrices. This parametric probability density function can be used to calculate the probability ($\text{Prob}(\max |d_{j,l_j,k_1,k_2}^{\text{noise}}| < |d_{j,l_j,k_1,k_2}^{\text{hist}}|)$) that the absolute value of a coefficient $|d_{j,l_j,k_1,k_2}^{\text{hist}}|$ from the Mexican hat contourlet transform of a two-dimensional histogram is less than or equal to the absolute value of the largest magnitude coefficient $\max |d_{j,l_j,k_1,k_2}^{\text{noise}}|$ from the Mexican hat contourlet transform of a random matrix produced by heteroskedastic noise as specified by the stochastic model in Section 8.2.6. This probability is incorporated into a ‘weight’ $w_{j,l_j,k_1,k_2} = \text{Prob}(\max |d_{j,l_j,k_1,k_2}^{\text{noise}}| < |d_{j,l_j,k_1,k_2}^{\text{hist}}|)$ that is calculated for every scale, direction and location index of the Mexican hat contourlet transform. Each coefficient d_{j,l_j,k_1,k_2} has its own unique weight. Furthermore, these weights can then be used for smoothing the two-dimensional histogram grid \mathbf{F} (the discrete analog of $f_{\text{raw}} [\Re(z_{l,a,x,y}^j), \Im(z_{l,a,x,y}^j)]$) by multiplying the coefficients d_{j,l_j,k_1,k_2} produced by the Mexican hat contourlet transform of the grid \mathbf{F} with the weights w_{j,l_j,k_1,k_2} . This smoothing method is in fact a special type of thresholding policy that modifies the coefficient magnitude d_{j,l_j,k_1,k_2} according to a function (that is based on $p(z, \mu, \sigma, \xi)$ rather than a thresholding policy such as that used by Do & Vertelli (2003)) that sets coefficients below a certain fixed magnitude to zero.

8.2.8.2 Observations on the Maximum Absolute Magnitude of Coefficients of the Mexican Hat Contourlet Transform of Random Matrices

In order to identify signal from noise using the coefficients of the Mexican hat contourlet transform, we need to understand how the coefficients respond to signals (or features) in the two-dimensional histograms. The Mexican hat contourlet transform is designed to capture clusters of data in the two-dimensional histograms. It achieves this by the use of scaling $\psi_{1,k_1,k_2}(x, y)$ and detail $\rho_{j,l_j,k_1,k_2}(x, y)$ functions that examine the two-dimensional histograms over a range of scales, directions and orientations. Large magnitude coefficients are produced when the scaling or detail functions capture many non-zero entries in the two-dimensional histogram or when the inner product between the scaling or detail functions and the two-dimensional histogram is large.

The two-dimensional histogram may exhibit non-zero entries that cluster close together (in location and magnitude of the frequencies) or that are aligned in a particular direction. These relationships between entries in the two-dimensional histogram will produce correlations between the coefficients of its Mexican hat contourlet transform. As such the Mexican hat contourlet transform coefficients of the two-dimensional histogram will have larger magnitude and stronger correlations when the scaling $\phi_{1,k_1,k_2}(x, y)$ and detail $\rho_{j,l_j,k_1,k_2}(x, y)$ functions closely match the scale, orientation and location of the underlying structure in the two-dimensional histogram.

In contrast, random data such as that specified by the **E** grid and the heteroskedastic noise model of Section 8.2.6 can not in general be expected to produce clusters of observations and hence large magnitude coefficients or correlations. Numerical experiments on simulated random data matrices (which are meant to emulate the **E** grid) are used to demonstrate that the corresponding Mexican hat contourlet transform coefficients are of relatively low-magnitude.

Matrices of random data (denoted $\mathbf{N}_{s,i}$, $i = 1, \dots, 10000$) were generated in order to perform simulation studies. The dimensions of each of the $\mathbf{N}_{s,i}$ matrices was 256 by 256 entries. The magnitude each entry $\mathbf{N}_{s,i}(u, v)$ ($u = 1, \dots, 256, v = 1, \dots, 256$) was obtained by independently sampling a $\mathcal{N}(0, \sigma_r^2)$, $\sigma_r^2 \sim U(0, 1)$ distribution and assigning the sampled value as the magnitude of random data matrix $\mathbf{N}_{s,i}(u, v)$ at the location (u, v) . The variance parameter σ_r^2 is re-sampled from a $U(0, 1)$ distribution using the Mersenne Twister algorithm (Matsumoto & Nishimura 1998) for each shift in the location index (u, v) . As indicated by the index i , a total of 10,000 such random matrices were generated and stored as the ‘simulated noise’ data set. An example from this data set is displayed in Figure 8.6(a), note the lack of any readily identifiable visual features in the matrix.

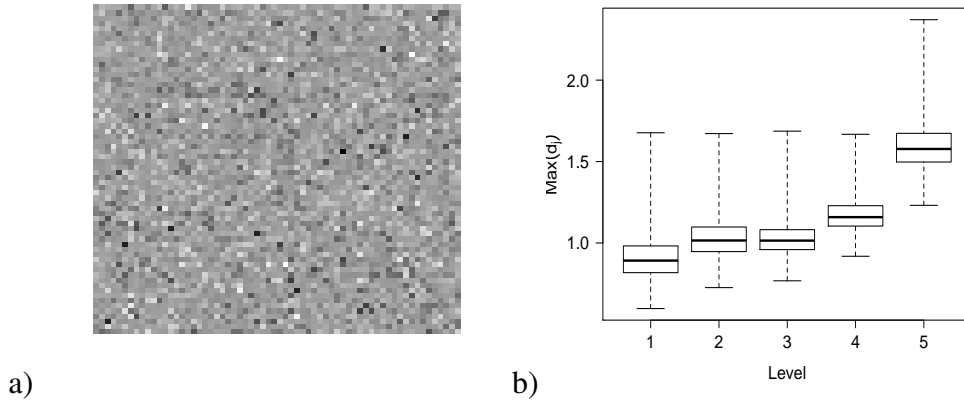


Figure 8.6: Simulation studies using the heteroskedastic noise model and the Mexican hat contourlet transform: (a) an example of simulated heteroskedastic noise (samples drawn from a $\mathcal{N}(0, \sigma_r^2)$ distribution where $\sigma_{i,j} \sim U(0, 1)$, drawn randomly for each of the $u = 1, \dots, 256$ and $v = 1, \dots, 256$ locations. The matrix is presented using a grey scale color map where darker shades correspond to lower coefficient magnitudes and (b) boxplots of maxima of the absolute values ($\text{Max}(|d_j|)$) of Mexican contourlet transform coefficients of heteroskedastic noise across all direction bands for scales $j = 1, \dots, 5$.

The Mexican hat contourlet transform (Section 8.2.4) was applied to each of the ten thousand $N_{s,i}$ matrices of random data. A total of $J = 5$ scales was used in the transform with $2^j, j = 1, 2, 3, 4, 5$ direction bands per scale. That is scale $j = 1$ has $l_1 = 2^1 = 2$ directions bands whilst scale $j = 5$ has $l_5 = 2^5 = 32$ direction bands. Therefore, this particular version of the Mexican hat contourlet transform is designed to double the number of direction bands for every increase in scale. At the finest scale, many (thirty two) different directions are analysed for ‘features’ in the matrices of random data.

The result of applying the Mexican hat contourlet transform (with the parameters $J = 5, l_j = 2^j$ to each matrix $N_{s,i}$ of random data is the set of coefficients $\{C_{1,i}, D_{1,1,i}, D_{1,2,i}, \dots, D_{5,32,i}\}$ where the index i indicates which out of the ten thousand matrices that the transform was performed on. After applying the Mexican hat contourlet transform (with parameters $J = 5, l_j = 2^j$) to all ten thousand random data matrices *super-arrays* of coefficients are produced.

These super-arrays are denoted $\{C_1, D_{1,1}, D_{1,2}, \dots, D_{5,32}\}$ and are formed by stacking each of the ten thousand matrices $\{C_{1,i}, D_{1,1,i}, D_{1,2,i}, \dots, D_{5,32,i}\}$ on top of each other along a depth dimension. Each of the super-arrays $\{C_1, D_{1,1}, D_{1,2}, \dots, D_{5,32}\}$ has three dimensions, the first corresponding to the row position (according to index $k_1 = 1, \dots, K_1$), the second corresponding to column position (according to index $k_2 = 1, \dots, K_2$) and the third corresponding to image number (according to index $i = 1, \dots, 10000$). These super-arrays encode the analysis of the Mexican hat contourlet transform for each matrix of random data, across each scale, direction and location.

In order to enable further comparisons, the magnitudes of the entries in the super-arrays are modified by taking the absolute value. Across the depth index i (corresponding to matrix number) for each of the modified super-arrays $\{C_1^*, D_{1,1}^*, D_{1,2}^*, \dots, D_{5,32}^*\}$ the maximum value is found and recorded at index (a, b) in a new matrix M . The row index a of matrix M has sixty three values that refer to a particular scale-direction super-array. For instance, the index $a = 1$ corresponds to the C_1 super-array, whilst index $a = 63$ corresponds to the $D_{5,32}$ super-array. The column index b of the matrix M has ten-thousand values and corresponds to the matrix of random data from which the coefficients were originally sourced.

Thus, the matrix \mathbf{M} is of the form,

$$\mathbf{M} = \begin{bmatrix} \max(\mathbf{C}_{1,1}^*) & \max(\mathbf{C}_{1,2}^*) & \dots & \max(\mathbf{C}_{1,10000}^*) \\ \max(\mathbf{D}_{1,1,1}^*) & \max(\mathbf{D}_{1,1,2}^*) & \dots & \max(\mathbf{D}_{1,1,10000}^*) \\ \vdots & \vdots & \vdots & \vdots \\ \max(\mathbf{D}_{5,32,1}^*) & \max(\mathbf{D}_{5,32,2}^*) & \dots & \max(\mathbf{D}_{5,32,10000}^*) \end{bmatrix} \quad (8.11)$$

The data are further reduced for graphical and numerical summary using another matrix \mathbf{M}^* . This matrix has five rows (which refer to scale j) and ten thousand columns (that refer to the i th matrix of random data). The entry in the j th row and i th column of the \mathbf{M}^* matrix is formed by taking the maximum of all of the entries from the \mathbf{M} matrix that are relevant to the detail coefficients at j th scale of the Mexican hat contourlet transform for the i th matrix of random data. For instance, $\mathbf{M}^*(1, i) = \max[\mathbf{M}(2, i), \mathbf{M}(3, i)] = \max[\max(\mathbf{D}_{1,1,i}^*), \max(\mathbf{D}_{1,2,i}^*)]$. Five boxplots are plotted in Figure 8.6(b), each was found by extracting the values for a specific row of the \mathbf{M}^* matrix for all ten thousand columns. These boxplots report on the distribution of the maximum absolute value of the coefficients at the j th scale (across all directions) for the Mexican hat contourlet transform of matrices of random data. Observe that the boxplots are ordered according to increasing scale parameter $j = 1, 2, 3, 4, 5$ and that in general the maximum absolute value of coefficient magnitude tends to increase with scale. Furthermore, the boxplots suggest that the absolute value of any coefficient in the Mexican hat contourlet transform of matrices of random data whose entries were formed by sampling $\mathcal{N}(0, \sigma_r^2)$, $\sigma_r^2 \sim U(0, 1)$ distributions is very unlikely to be above a magnitude of three. The ranges of these maximum absolute coefficient magnitudes are reported in first column of Table 8.1 on page 228, they will be compared to the ranges of the maximum absolute coefficient magnitudes from the Mexican hat contourlet transforms of sampled bivariate probability density functions in Section 8.2.8.3.

The sampled bivariate probability density functions will be generated by creating a discrete grid \mathbf{S} of locations $(s, t) \in \mathbb{R}_+^2$, $s = 1, \dots, S$, $t = 1, \dots, T$, calculating the values of a pre-assigned probability density function (denoted $f(x, y)$) for each of the grid locations ($x = s, y = t$). The result $f(x = s, y = t)$ is assigned to entry $\mathbf{S}(s, t)$ in the discrete grid. This is repeated for all $s = 1, \dots, S$, $t = 1, \dots, T$ locations in the grid.

Comparison of the coefficients from the Mexican hat contourlet transform of random matrices to the coefficients from the Mexican hat contourlet transform of matrices of sampled bivariate probability density functions will be very informative and indicate substantial differences in the magnitude of the coefficients corresponding to these two different types of data. These differences will motivate a method in Section 8.2.8.4 that smooths the two-dimensional histograms of Section 8.2.5 using the Mexican hat contourlet transform by modifying the magnitude of the Mexican hat contourlet transform coefficients.

$$\max(|d_j|)$$

Scale	Random	Density
1	[0.60, 1.68]	[0.50, 20.92]
2	[0.73, 1.67]	[2.72, 54.37]
3	[0.77, 1.69]	[7.11, 81.16]
4	[0.92, 1.67]	[12.62, 115.66]
5	[1.23, 2.37]	[20.71, 165.14]

Table 8.1: Maximum absolute magnitude $|d_j|$ of the Mexican hat contourlet transform coefficients (for $J = 5$ scales) across direction and location for both simulated random and sampled bivariate probability density function data. Values reported to two decimal places.

8.2.8.3 Observations on the Maximum Absolute Magnitude of the Coefficients from the Mexican Contourlet Transform of Bivariate Density Functions

This section compares the range of the absolute values of the coefficients (across scales) of the Mexican hat contourlet transform of simulated random data (Section 8.2.8.2) to the range of the absolute values of the coefficients (across scales) of the Mexican hat contourlet transform of matrices that were sampled from bivariate probability density functions. The results will demonstrate that range of the Mexican hat contourlet transform coefficients of the simulated random data and the range of the Mexican hat contourlet transform coefficients produced by the matrices of bivariate probability density functions does not overlap at the higher ($j = 2, \dots, 5$) scales. From these results it will be hypothesised that the Mexican hat contourlet transform coefficients of the simulated random data at scales ($j = 2, \dots, 5$) is of a substantially lower magnitude than the Mexican hat contourlet transform coefficients produced by the matrices of a wide range of bivariate probability density functions. This hypothesis will assist in developing a probability density function estimation method based upon the Mexican hat contourlet transform in Section 8.2.8.4. This probability density function estimation method weights the Mexican hat contourlet transform coefficients of a two-dimensional histogram according to their probability of being produced by ‘noise’ (random data) which is in turn estimated using a separate probability density function model of the absolute value of the Mexican hat contourlet transform coefficients.

The focus of this section is on those probability density functions that do not resemble bivariate Gaussian probability density functions. Such probability density functions are termed ‘non-Gaussian’. Probability density functions that are non-Gaussian exhibit features such as multiple local maxima or a mixture of different frequency components. The XOR estimation model of Section 8.2.2 would (in most cases) indicate that a non-parametric representation (which is based upon the Mexican hat contourlet transform) be used to describe those probability density functions (of the adaptive image transform coefficients) which are non-Gaussian.

In order to compare the Mexican hat contourlet transform coefficients produced by random data to those of Mexican hat contourlet transform coefficients produced from samples of known probability density functions, we must first prepare a database of probability density function standards to which the Mexican hat contourlet transform can be applied. The Marron and Wand (1992) family of probability density functions was selected as the database of probability density functions standards. The family of probability density functions created by Marron and Wand

(1992) consists of fourteen different non-Gaussian probability density functions, as well as the standard Gaussian probability density function. All of the non-Gaussian functions in the Marron and Wand (1992) family of probability density functions are constructed as a linear combination of Gaussian probability density functions. Marron and Wand (1992) demonstrated that a wide range of probability density functions could be represented as a linear combination of such Gaussian probability density functions.

The probability density functions that were specified in table 1, page 720 of Marron and Wand (1992) were one-dimensional. We extended these probability density functions to two-dimensions to allow the Mexican hat contourlet transform to be applied to them.

The specific probability density functions included in the database of probability density functions standards were:

- (i) Standard Bivariate Gaussian: $p_1(x, y) = \mathcal{N}(0, 0, 1, 1, 0)$.

In other words a bivariate Gaussian probability density function with $\mu_x = 0$, $\mu_y = 0$, $\sigma_x^2 = 1$, $\sigma_y^2 = 1$, $\rho_{x,y} = 0$ where $\rho_{x,y}$ is the correlation coefficient of the probability density function.

- (ii) Skewed Unimodal:

$$p_2(x, y) = \frac{1}{5}\mathcal{N}(0, 0, 1, 1, 0) + \frac{1}{5}\mathcal{N}(\frac{1}{2}, \frac{1}{2}, (\frac{2}{3})^2, (\frac{2}{3})^2, 0) + \frac{3}{5}\mathcal{N}(\frac{13}{12}, \frac{13}{12}, (\frac{5}{9})^2, (\frac{5}{9})^2, 0).$$

- (iii) Strongly Skewed:

$$p_3(x, y) = \sum_{l=0}^7 \frac{1}{8}\mathcal{N}(3((\frac{2}{3})^l - 1), 3((\frac{2}{3})^l - 1), (\frac{2}{3})^{2l}, (\frac{2}{3})^{2l}, 0).$$

- (iv) Kurtotic Unimodal:

$$p_4(x, y) = \frac{2}{3}\mathcal{N}(0, 0, 1, 1, 0) + \frac{1}{3}\mathcal{N}(0, 0, (\frac{1}{10})^2, (\frac{1}{10})^2, 0).$$

- (v) Outlier:

$$p_5(x, y) = \frac{1}{10}\mathcal{N}(0, 0, 1, 1, 0) + \frac{9}{10}\mathcal{N}(0, 0, (\frac{1}{10})^2, (\frac{1}{10})^2, 0).$$

- (vi) Bimodal:

$$p_6(x, y) = \frac{1}{2}\mathcal{N}(-1, -1, (\frac{2}{3})^2, (\frac{2}{3})^2, 0) + \frac{1}{2}\mathcal{N}(1, 1, (\frac{2}{3})^2, (\frac{2}{3})^2, 0).$$

- (vii) Separated Bimodal:

$$p_7(x, y) = \frac{1}{2}\mathcal{N}(-\frac{3}{2}, -\frac{3}{2}, (\frac{1}{2})^2, (\frac{1}{2})^2, 0) + \frac{1}{2}\mathcal{N}(\frac{3}{2}, \frac{3}{2}, (\frac{1}{2})^2, (\frac{1}{2})^2, 0).$$

- (viii) Skewed Bimodal:

$$p_8(x, y) = \frac{3}{4}\mathcal{N}(0, 0, 1, 1, 0) + \frac{1}{4}\mathcal{N}(\frac{3}{2}, \frac{3}{2}, (\frac{1}{3})^2, (\frac{1}{3})^2, 0).$$

(ix) Trimodal:

$$p_9(x, y) = \frac{9}{20}\mathcal{N}(-\frac{6}{5}, -\frac{6}{5}, (\frac{3}{5})^2, (\frac{3}{5})^2, 0) + \frac{9}{20}\mathcal{N}(\frac{6}{5}, \frac{6}{5}, (\frac{3}{5})^2, (\frac{3}{5})^2, 0) + \frac{1}{10}\mathcal{N}(0, 0, (\frac{1}{4})^2, (\frac{1}{4})^2, 0).$$

(x) Claw:

$$p_{10}(x, y) = \frac{1}{2}\mathcal{N}(0, 0, 1, 1, 0) + \sum_{l=0}^4 \frac{1}{10}\mathcal{N}(\frac{l}{2} - 1, \frac{l}{2} - 1, (\frac{1}{2})^2, (\frac{1}{2})^2, 0).$$

(xi) Double Claw:

$$p_{11}(x, y) = \frac{49}{100}\mathcal{N}(-1, -1, (\frac{2}{3})^2, (\frac{2}{3})^2, 0) + \frac{49}{100}\mathcal{N}(1, 1, (\frac{2}{3})^2, (\frac{2}{3})^2, 0) + \sum_{l=0}^6 \frac{1}{350}\mathcal{N}(\frac{(l-3)}{2}, \frac{(l-3)}{2}, (\frac{1}{100})^2, (\frac{1}{100})^2, 0).$$

(xii) Asymmetric Claw:

$$p_{12}(x, y) = \frac{1}{2}\mathcal{N}(0, 0, 1, 1, 0) + \sum_{l=-2}^2 (\frac{2^{1-l}}{31})\mathcal{N}(l + \frac{1}{2}, l + \frac{1}{2}, (\frac{2^{-l}}{10})^2, (\frac{2^{-l}}{10})^2, 0).$$

(xiii) Asymmetric Double Claw:

$$p_{13}(x, y) = \sum_{l=0}^1 \frac{46}{100}\mathcal{N}(2l-1, 2l-1, (\frac{2}{3})^2, (\frac{2}{3})^2, 0) + \sum_{l=1}^3 \frac{1}{300}\mathcal{N}(-\frac{l}{2}, -\frac{l}{2}, (\frac{1}{100})^2, (\frac{1}{100})^2, 0) + \sum_{l=1}^3 \frac{7}{300}\mathcal{N}(\frac{l}{2}, \frac{l}{2}, (\frac{7}{100})^2, (\frac{7}{100})^2, 0).$$

(xiv) Smooth Comb:

$$p_{14}(x, y) = \sum_{l=0}^5 (\frac{2^{5-l}}{63})\mathcal{N}(\frac{[65-96(\frac{1}{2})^l]}{21}, \frac{[65-96(\frac{1}{2})^l]}{21}, (\frac{32}{63})^2, (\frac{32}{63})^2, 0).$$

(xv) Discrete Comb:

$$p_{15}(x, y) = \sum_{l=0}^2 \frac{2}{7}\mathcal{N}(\frac{(12l-15)}{7}, \frac{(12l-15)}{7}, (\frac{2}{7})^2, (\frac{2}{7})^2, 0) + \sum_{l=8}^{10} \frac{1}{21}\mathcal{N}(\frac{2l}{7}, \frac{2l}{7}, (\frac{1}{21})^2, (\frac{1}{21})^2, 0).$$

(adapted from table 1, page 720, Marron & Wand 1992).

Discrete representations of each of the above fifteen probability density functions were obtained by creating a discrete grid of $x = (1, \dots, N_x)$ and $y = (1, \dots, N_y)$ values. At each location (x^*, y^*) in this grid the value of $p_r(x^*, y^*)$ of the r th out of the $r = 1, \dots, 15$ probability density functions in the modified Marron & Wand (1992) database was calculated. The result was a set of fifteen matrices \mathbf{P}_r ($r = 1, \dots, 15$) that provided a discrete matrix representation of the probability density functions in the modified Marron and Wand (1992) database that were amenable to input into the Mexican hat contourlet transform. Examples of the separated bimodal P_7 and the trimodal P_9 probability density functions are displayed in Figures 8.7(a) and (b) respectively. Note that the bivariate Gaussian probability density function $p_1(x, y)$ was included in the study in order to simulate a situation in which the XOR estimation model erroneously specifies a non-parametric model be fit to a bivariate Gaussian probability density function.

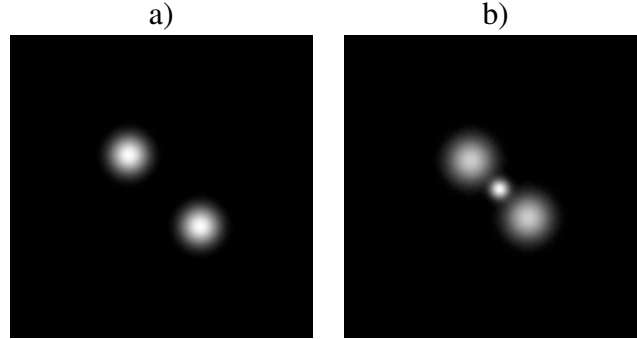


Figure 8.7: Examples of functions in the Marron & Wand (1992) database: (a) the separated bimodal and (b) trimodal probability density functions.

The Mexican hat contourlet transform was applied to each of the fifteen P_r matrices that represent the probability density functions in the modified Marron & Wand (1992) database. The maximum of the absolute values of the coefficients for each scale $j \in J$ across all directions of each Mexican hat contourlet transform was recorded. The result is a vector $\mathbf{m}_r = [\max(|\mathbf{C}|_1, r, \max |D|_{1,r}, \dots, \max |D|_{J,r})]$ for the coefficients corresponding to Mexican hat contourlet transform of the P_r th matrix. These vectors are themselves stored in another matrix \mathbf{M}_p , which is defined as,

$$\mathbf{M}_p = \begin{bmatrix} \max(|\mathbf{C}_{1,1}|) & \max(|\mathbf{D}_{1,1}|) & \dots & \max(|\mathbf{D}_{J,1}|) \\ \max(|\mathbf{C}_{1,2}|) & \max(|\mathbf{D}_{1,2}|) & \dots & \max(|\mathbf{D}_{J,2}|) \\ \vdots & \vdots & \vdots & \vdots \\ \max(|\mathbf{C}_{1,15}|) & \max(|\mathbf{D}_{1,15}|) & \dots & \max(|\mathbf{D}_{J,15}|) \end{bmatrix}. \quad (8.12)$$

The entries of the M_p matrix are reported in Table 8.2 but with specific reference to each probability density function in the modified Marron & Wand (1992) database. From inspection of Table 8.2 it is evident that the maximum absolute values of coefficient magnitudes vary greatly both between scales for the same density and across a scale between densities. Of central importance is that in general, the maximum absolute magnitude of the Mexican hat contourlet transform of the probability density function data appears to be of larger than the maximum absolute magnitude of the Mexican hat contourlet transform of random noise. The second column of Table 8.1 provides the range of these values for each scale in the Mexican hat contourlet transform. Observe that in general the maximum absolute magnitude of Mexican hat contourlet transform coefficients for probability density functions is much greater than the corresponding maximum

Pdf	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5
$p_1(x, y)$	0.50	2.85	9.10	20.64	41.01
$p_2(x, y)$	1.02	5.74	17.42	36.03	60.66
$p_3(x, y)$	0.94	4.20	9.56	17.32	29.00
$p_4(x, y)$	20.92	54.37	81.16	115.66	165.14
$p_5(x, y)$	20.64	52.74	76.09	103.96	142.21
$p_6(x, y)$	0.57	3.21	9.97	21.50	41.40
$p_7(x, y)$	1.02	5.66	17.00	34.37	64.08
$p_8(x, y)$	1.06	5.91	17.60	35.37	58.75
$p_9(x, y)$	0.88	4.34	10.84	22.45	42.49
$p_{10}(x, y)$	6.43	18.87	28.09	50.69	68.34
$p_{11}(x, y)$	8.40	12.97	28.38	59.87	112.92
$p_{12}(x, y)$	8.07	14.11	33.16	63.20	106.44
$p_{13}(x, y)$	9.65	15.21	29.13	59.81	112.64
$p_{14}(x, y)$	0.58	2.85	8.59	17.47	29.13
$p_{15}(x, y)$	0.58	2.72	7.11	12.62	20.71

Table 8.2: Maximum of the absolute values of coefficient magnitude for the Mexican hat contourlet transform of each probability density function in the modified Marron & Wand (1992) database.

absolute magnitude of Mexican hat contourlet transform coefficients for random noise. Furthermore the ranges of the maximum absolute magnitude of the coefficients for the noise and density data does not overlap for scales $j = 2$ to 5. These results suggest substantial differences between the magnitudes of coefficients of the Mexican hat contourlet transform of random data and the magnitudes of coefficients of the Mexican hat contourlet transform of probability density functions. A Mexican hat coefficient weighting or thresholding strategy will be formulated in Section 8.2.8.4 based upon these observed differences.

8.2.8.4 A Thresholding Policy for Mexican Hat Contourlet Transform Coefficients based on the Generalised Extreme Value Distribution

A thresholding policy for the coefficients of the Mexican hat contourlet transform of the two-dimensional histogram, denoted by the set $\{d_{j,l_j,k_1,k_2}^{\text{hist}}\}$, is performed by coefficient weighting. The weights are derived by comparing the magnitude of $|d_{j,l_j,k_1,k_2}^{\text{hist}}|$ with a distribution of coefficients obtained from noise, $|d_{j,l_j,k_1,k_2}^{\text{noise}}|$.

Where a coefficient $|d_{j,l_j,k_1,k_2}^{\text{hist}}|$ is unlikely to have arisen from noise, it is given a large weight and vice-versa. The distribution derived from the coefficients $|d_{j,l_j,k_1,k_2}^{\text{noise}}|$ was taken to be a generalised extreme value distribution $P(|d_{j,l_j,k_1,k_2}^{\text{hist}}|, \mu, \sigma, \xi)$ for reasons explained below.

The weight w_{j,l_j,k_1,k_2} for each coefficient $|d_{j,l_j,k_1,k_2}^{\text{hist}}|$ is the probability that a value less than $|d_{j,l_j,k_1,k_2}^{\text{hist}}|$ could arise by a noise process. An example of the weighting function that is used to calculate such weights is displayed in Figure 8.8.

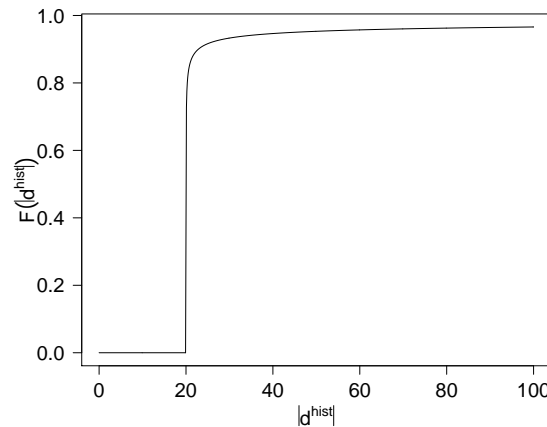


Figure 8.8: An example of a coefficient weighting function based upon the generalised extreme value distribution $P(|d_{j,l_j,k_1,k_2}^{\text{hist}}|, 20, 0.01, 3)$. Note how the lower part of the function (for $|d^{\text{hist}}| < 20$) resembles a step function, whilst the upper part $|d^{\text{hist}}| > 20$ resembles a smooth curve. This coefficient weighting function can be considered a combination of a ‘hard’ (set all coefficients below a certain magnitude to zero) and a ‘soft’ (down-weight smaller coefficient magnitudes) thresholding functions.

The weighted coefficients are then,

$$d_{j,l_j,k_1,k_2}^* = w_{j,l_j,k_1,k_2} \times d_{j,l_j,k_1,k_2} \quad (8.13)$$

where $w_{j,l_j,k_1,k_2} = \text{Prob}_{\hat{P}}(|d_{j,l_j,k_1,k_2}^{\text{noise}}| < |d_{j,l_j,k_1,k_2}^{\text{hist}}|)$

The weights allocated to the coefficients $|d_{j,l_j,k_1,k_2}^{\text{hist}}|$ are derived from a probability distribution function of the noise which is determined by simulation. Where $|d_{j,l_j,k_1,k_2}^{\text{hist}}|$ are coefficients arising from informative features, they are larger than those which emanate from noise and so coefficients will occur in the extreme right tail of the distribution function. If a distribution function was estimated using all of the coefficients pertaining to noise then the tail of the distribution function will be sparsely modelled because large coefficients arising from noise will be extremely rare.

An alternative method for determining the weights w_{j,l_j,k_1,k_2} needs to be found. We consider that a more suitable probability density function from which to calculate the weights is $\hat{P}(\max |d_{j,l_j,k_1,k_2}^{\text{noise}}|)$. This is generated with bootstrap methodology by simulating $\{|d_{j,l_j,k_1,k_2}^{\text{noise}}|\}$ many times and saving the maximum of each simulation (Efron & Tibshirani 1993). These maxima are summarised as a generalised extreme value distribution.

Fisher & Tippett (1928), Gnedenko (1943) and Gumbel (1958) provided the key statistical foundations of extreme value theory. Extreme value theory concerns the asymptotic behaviour of the distribution of the maximum values of sequences as the number of entries in each sequence approaches infinity. The theory of extreme values can be applied in two-dimensional histogram smoothing to find, $\text{Prob}_{\hat{P}}(|\max d_{j,l_j,k_1,k_2}^{\text{noise}}| < |d_{j,l_j,k_1,k_2}^{\text{hist}}|)$.

Von Mises (1954) and Jenkinson (1955) described the generalised extreme value probability distribution function $P(z, \mu, \sigma, \xi)$. This probability distribution function is defined as,

$$P(z, \mu, \sigma, \xi) = \exp \left\{ - \left[1 + \left(\frac{z - \mu}{\sigma} \right) \times \xi \right]^{-\frac{1}{\xi}} \right\} \quad (8.14)$$

for a random variable $z \in \mathbb{R}$, a location parameter $\mu \in \mathbb{R}$, scale parameter $\sigma \in \mathbb{R}^+$, shape parameter $\xi \in \mathbb{R}$ but with the constraint $[z : 1 + \xi(z - \mu)/\sigma > 0]$ (equation 1.4, page 4, Kotz & Nadarajah 2000).

The simulations that are used to find the maximum of the Mexican hat contourlet transform of noise proceeds in the following manner:

- i) A total of n random matrices ($n > 100$) are generated using the methods of Section 8.2.8.2.
- ii) Each random matrix is analysed using the Mexican hat contourlet transform.
- iii) The maximum absolute coefficient magnitude for each scale j and direction l_j in the coefficient matrix \mathbf{D}_{j,l_j} is found.
- iv) These maximum values are recorded in a vector \mathbf{m}_{j,l_j} in which the i th out of the n entries encodes the maximum absolute coefficient magnitude for the matrix \mathbf{D}_{j,l_j} of the Mexican hat contourlet transform of the i th matrix of random data.
- v) The maximum absolute values of all the coefficients at each scale $j = 1, \dots, J$ for each of the n transforms is then recorded. That is, for scale j we create the vector

$$\mathbf{m}_j = [\max(\max \mathbf{D}_{1,1}^{(1)}, \dots, \max \mathbf{D}_{1,2^j}^{(1)}), \dots, [\max(\max \mathbf{D}_{1,1}^{(n)}, \dots, \max \mathbf{D}_{1,2^j}^{(n)})]]. \quad (8.15)$$

The problem for estimating $\hat{P}(\max |d_{j,l_j,k_1,k_2}^{\text{noise}}|)$ is about the same as that for estimating $\hat{P}(|d_{j,l_j,k_1,k_2}^{\text{noise}}|)$. There are insufficient numbers of maximum values of coefficients, $\max |d_{j,l_j,k_1,k_2}^{\text{noise}}|$, that are of comparable magnitude to many of the coefficients $|d_{j,l_j,k_1,k_2}^{\text{hist}}|$ that will be encountered in the smoothing of two-dimensional histograms. This problem will still occur even when the number of coefficients in the sequence from which $\max |d_{j,l_j,k_1,k_2}^{\text{noise}}|$ is found is very large ($n \rightarrow \infty$), as very large magnitudes of $|d_{j,l_j,k_1,k_2}^{\text{noise}}|$ are rare and extreme events.

The generalised extreme value probability distribution function $P(z, \mu, \sigma, \xi)$ however, can be used to describe the probability that the maximum value in a large sequence of coefficients will not be greater than a certain fixed value. It can be used to overcome the probability distribution function estimation problem and be used to find the coefficient weights as,

$$\text{Prob}_{\hat{P}}(|\max d_{j,l_j,k_1,k_2}^{\text{noise}}| < |d_{j,l_j,k_1,k_2}^{\text{hist}}| \approx \hat{P}(z = |d_{j,l_j,k_1,k_2}^{\text{hist}}|, \hat{\mu}, \hat{\sigma}, \hat{\xi}).$$

A generalised extreme value distribution function is fit for every scale of the Mexican hat contourlet transform to produce a family of coefficient weighting functions

$[\hat{P}(z, \hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1), \hat{P}(z, \hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2), \dots, \hat{P}(z, \hat{\mu}_J, \hat{\sigma}_J, \hat{\xi}_J)]$. In each instance, the parameters (μ_j, σ_j, ξ_j) for $j = 1, \dots, J$ of each generalised extreme value probability distribution function $P(z, \mu_j, \sigma_j, \xi_j)$ were estimated using algorithm AS 215 (Hosking 1985). This algorithm uses iterative methods (a modified version of Newton-Raphson iteration) to produce maximum likelihood estimates $(\hat{\mu}_j, \hat{\sigma}_j, \hat{\xi}_j)$ of the parameters (μ_j, σ_j, ξ_j) in conjunction with a data set (in this case the \mathbf{m}_j vector) which is assumed to consist of samples all drawn from the same generalised extreme value distribution function $P(z, \mu_j, \sigma_j, \xi_j)$.

The estimates $[\hat{P}(z, \hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1), \hat{P}(z, \hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2), \dots, \hat{P}(z, \hat{\mu}_J, \hat{\sigma}_J, \hat{\xi}_J)]$ encode the information needed to obtain the weights w_{j,l_j,k_1,k_2} that are used to obtain the modified Mexican hat contourlet transform coefficients $d_{j,l_j,k_1,k_2}^{*\text{ hist}}$, which in turn allows estimation of the smooth probability density function, $\tilde{\mathbf{F}}_{N\text{adj}}^g$.

Modification of the coefficients of the Mexican hat contourlet transform of the two dimensional histogram $\mathbf{F}_{g_{N\text{adj}}}$ is achieved as follows:

- i) Obtain the absolute values of the reference coefficients $d_{j,l_j,k_1,k_2}^{\text{hist}}$ of the Mexican hat contourlet transform of the two-dimensional histogram, $\mathbf{F}_{g_{N\text{adj}}}$ that is of interest. In other words calculate, $|d_{j,l_j,k_1,k_2}^{\text{hist}}|$.
- ii) For the reference coefficient $d_{j,l_j,k_1,k_2}^{\text{hist}}$ of the Mexican hat contourlet transform of the two-dimensional histogram $\mathbf{F}_{g_{N\text{adj}}}$, calculate $\text{Prob}(\max |d_{j,l_j,k_1,k_2}^{\text{noise}}| \leq |d_{j,l_j,k_1,k_2}^{\text{hist}}|)$ using the generalised extreme value probability distribution function estimate $\hat{P}(z, \hat{\mu}_j, \hat{\sigma}_j, \hat{\xi}_j)$ that has the same scale parameter index (j), as the reference coefficient $d_{j,l_j,k_1,k_2}^{\text{hist}}$.
- iii) Modify the coefficient $d_{j,l_j,k_1,k_2}^{\text{hist}}$ to be $d_{j,l_j,k_1,k_2}^{*,\text{hist}} = \text{Prob}(\max |d_{j,l_j,k_1,k_2}^{\text{noise}}| \leq |d_{j,l_j,k_1,k_2}^{\text{hist}}|) \times d_{j,l_j,k_1,k_2}^{\text{hist}}$.
- iv) Do steps (i)-(iii) for all coefficients $d_{j,l_j,k_1,k_2}^{\text{hist}}$ in the Mexican hat contourlet transform of the two-dimensional histogram $\mathbf{F}_{g_{N\text{adj}}}$, to obtain the set of modified coefficient matrices $(\mathbf{D}_{1,1}^*, \mathbf{D}_{1,2}^*, \dots, \mathbf{D}_{J,2^J}^*)$.

The smooth probability density function estimate $\tilde{F}_{g_{N\text{adj}}}$ (of the real and imaginary components of the adaptive image transform coefficients for fixed filter, scale, location and group), is then obtained by applying the inverse Mexican hat contourlet transform to the set of modified coefficient matrices $(\mathbf{C}_1^{\text{smooth}}, \mathbf{D}_{1,1}^*, \mathbf{D}_{1,2}^*, \mathbf{D}_{2,1}^*, \dots, \mathbf{D}_{J,2^J}^*)$ where the approximation coefficients $\mathbf{C}_1^{\text{smooth}}$ are the same as those of $\mathbf{C}_1^{\text{hist}}$ in the Mexican hat contourlet transform of the two-dimensional histogram, $\mathbf{F}_{g_{N\text{adj}}}$. In this manner, smooth probability density function estimates, $\tilde{\mathbf{F}}_{g_1}, \tilde{\mathbf{F}}_{g_2}, \dots, \tilde{\mathbf{F}}_{g_n}$ (in this case $\tilde{\mathbf{F}}_{g_{\text{norm}}}, \tilde{\mathbf{F}}_{g_{\text{ben}}}, \tilde{\mathbf{F}}_{g_{\text{mal}}}$) are obtained for all groups using the Mexican hat contourlet transform.

The smooth probability density function estimation process is repeated many times to obtain probability density estimates for each group corresponding to each filter, scale and location of the adaptive image transform coefficients that were retained by the filter function selection algorithm in Section 7.4.

8.2.9 Producing Probability Arrays for each SAXS Image

The methodology of Sections 8.2.1 to 8.2.8.4 can be combined into a cohesive framework to estimate the posterior probability of the tissue sample belonging to the normal, benign or malignant groups using the information provided by each adaptive image transform coefficient (at fixed filter, scale and location) independently. The probability density functions, $p_{g_{\text{norm}}}(z_{l,a,x,y})$, $p_{g_{\text{ben}}}(z_{l,a,x,y})$, $p_{g_{\text{mal}}}(z_{l,a,x,y})$ for the normal, benign and malignant groups were estimated using the adaptive image transform coefficients $z_{l,a,x,y}$ at fixed filter (l), scale (a) and location (x, y) indices for the SAXS images of a training data set. These probability density functions were estimated either parametrically (using a bivariate Gaussian distribution with maximum likelihood estimators) or non-parametrically (using the Mexican hat contourlet transform of the two-dimensional histogram smoother) based upon the outcome of separate Shapiro-Wilk tests on the real $\Re(z_{l,a,x,y})$ and the imaginary $\Im(z_{l,a,x,y})$ of the adaptive image transform coefficients of samples within each group. Probability density function estimates were obtained for the three groups for each filter, scale and location index of the adaptive image transform. The probability density functions of those filter, scale and location indices that were retained post the filter function selection algorithm (Section 7.4) were stored in computer memory as a component of the SAXS image diagnostic model.

Probability estimates ($\hat{\text{Prob}}_n(z_{l,a,x,y}^j)$, $\hat{\text{Prob}}_b(z_{l,a,x,y}^j)$, $\hat{\text{Prob}}_m(z_{l,a,x,y}^j)$) were then obtained for fixed coefficient indices (l, a, x, y) for the j th ($j = 1, \dots, N_{\text{val}}$) SAXS image in the *validation* data set by using the probability density function estimates, $\hat{p}_{g_{\text{norm}}}(z_{l,a,x,y})$, $\hat{p}_{g_{\text{ben}}}(z_{l,a,x,y})$, $\hat{p}_{g_{\text{mal}}}(z_{l,a,x,y})$ in Bayes' rule of equation (8.2). This process is repeated, one at a time, for all coefficient indices (l, a, x, y) of the adaptive image transform (post filter function selection) of the SAXS image.

The probability estimates ($\hat{\text{Prob}}_n(z_{l,a,x,y}^j)$, $\hat{\text{Prob}}_b(z_{l,a,x,y}^j)$, $\hat{\text{Prob}}_m(z_{l,a,x,y}^j)$) are stored in three arrays $\mathbf{P}_{j,\text{norm}}$, $\mathbf{P}_{j,\text{ben}}$ and $\mathbf{P}_{j,\text{mal}}$ that has the same dimensions as the adaptive image transform coefficient array \mathcal{D}^* (post application of the filter function selection algorithm). Thus the entry in the $(l, a, x, y) = (1, 3, 4, 5)$ in the $\mathbf{P}_{2,\text{norm}}$ array corresponds to the probability $\hat{\text{Prob}}_n(z_{1,2,3,4,5})$ for the normal group of the second SAXS image in the validation data set given the first filter, third scale and the location $(x, y) = (4, 5)$. The above process is repeated for all $j = 1, \dots, N_{\text{val}}$ SAXS images in the validation data set with three probability arrays $\mathbf{P}_{j,\text{norm}}$, $\mathbf{P}_{j,\text{ben}}$ and $\mathbf{P}_{j,\text{mal}}$ being created for each SAXS image.

8.2.10 Summary of Probability Density Function Estimation Using the Mexican Hat Contourlet Transform

The objective of this section was to estimate the posterior probability that a particular tissue sample belongs to a certain group (either normal, benign or malignant) based upon the magnitude of only one adaptive image transform coefficient $z_{l,a,x,y}^j$ with fixed filter, scale and location (l, a, x, y) indices. These posterior probabilities are estimated for all coefficients of the adaptive image transform for all groups in an independent manner. The end result is three (one each for the normal, benign and malignant groups) arrays of posterior probability estimates that have the same dimension as the coefficient array of the adaptive image transform of the SAXS image under investigation, post the filter function selection algorithm of Section 7.4.

Recapping, the methodology of this section was centered around the application of Bayes' rule in equation (8.2). Implementation of Bayes' rule required the estimation of three probability density functions, $p_{g_{\text{norm}}}(z_{l,a,x,y})$, $p_{g_{\text{ben}}}(z_{l,a,x,y})$, $p_{g_{\text{mal}}}(z_{l,a,x,y})$ for the normal, benign and malignant tissue groups for the magnitude of the coefficients $z_{l,a,x,y}$ of the adaptive image transform at fixed indices (l, a, x, y) . The result is that a potentially very large number of probability density functions need to be both estimated and stored. In order to reduce the computational and storage requirements of the model, bivariate Gaussian probability density functions were used to model the distribution of coefficient magnitude at those coefficient filter, scale and location indices for which the amorphous scatter feature of the SAXS image was captured. The probability density functions that describe the magnitude of the coefficient at the remaining indices of the adaptive image transform (post filter-function selection) were estimated non-parametrically using the Mexican hat contourlet transform. A raw two-dimensional histogram was constructed using a grid based upon the magnitudes of the real and imaginary components of the adaptive image transform coefficients in the training SAXS image data set at fixed group, filter, scale and location. The coefficients of the Mexican hat contourlet transform of this two-dimensional histogram were then modified by multiplying their magnitude with the probability that their absolute magnitude was less than or equal to the absolute magnitude of the coefficients of the Mexican hat contourlet transform of random heteroskedastic noise matrices. These probabilities were determined at each of the J scales of the Mexican hat contourlet transforms using a generalised extreme value distribution fit using the absolute magnitude of coefficients of the Mexican hat contourlet transform of simulated random heteroskedastic noise matrices. Once the coefficients of the Mexican hat contourlet transform of the two-dimensional histogram were modified,

they served as non-parametric estimates $\hat{p}_{g_{N\text{adj}}}(z_{l,a,x,y})$ of the probability density function of the group $g_{N\text{adj}}$ (either normal, benign or malignant) at filter, scale and location index (l, a, x, y) . This process was repeated for all groups and for all filter, scales and location indices of the adaptive image transform for which a bivariate Gaussian probability density function did not suffice. The select probability density function estimates (for all groups and all filter, scale and location indices that remained post filter function selection) were stored in computer memory. The posterior probability of each group could then be estimated for each coefficient in the adaptive image transform (post filter function selection) of each SAXS image in the validation data set using the relevant stored probability density function estimates in conjunction with Bayes' rule. For each SAXS image, these posterior probability estimates are stored in arrays $\mathbf{P}_{j,\text{norm}}$, $\mathbf{P}_{j,\text{ben}}$ and $\mathbf{P}_{j,\text{mal}}$ that are the same size as the original adaptive image transform (post filter function selection) coefficient arrays.

Posterior probability estimates can be obtained for each group at each filter, scale and location index of the coefficients of the adaptive image transform of a SAXS image. This potentially large number of probability estimates allows the extraction of the slightest piece of evidence that the tissue sample under investigation is malignant. Furthermore, it is possible to identify the scale and location of the intensity magnitude feature in the original SAXS image that indicates malignancy. The results of such an analysis might assist in furthering the scientific understanding of how the structure of collagen is altered in malignant breast tissue conditions.

If the SAXS imaging technology is eventually deployed in a histopathology laboratory, then the posterior probability arrays output by the methodology of this section can be inspected by a pathologist using a graphical user-interface to assist with the assessment of the tissue. Despite the wealth of information that can be gleaned by inspecting the posterior probability arrays for each image, there might be a desire by the pathologist to have a simplified overall probability that the tissue is normal, benign or malignant. An overall probability summary may also be useful in semi-automated screening applications of the SAXS imaging device. A tissue sample could be imaged using the SAXS camera and analysed using software based upon the methodology of this section. Tissue samples could be prioritised for inspection by a pathologist and the SAXS imaging system could therefore potentially reduce the time until diagnosis of the most urgent cases. The methodology that is used to estimate this overall probability of the sample being normal, benign or malignant breast is the subject of the next section.

8.3 Overall Probability Estimates Using Coefficient Information Across Scales

8.3.1 Overview

The objective of this section is to estimate the overall probability of a SAXS image being produced by either normal, benign or malignant breast tissue using the posterior probability arrays produced in Section 8.2. This section is divided into six components. This sub-section, Section 8.3.1 begins with a broad overview of the methodology that is developed.

Section 8.3.2 reviews the structure of the arrays $\mathbf{P}_{j,\text{norm}}$, $\mathbf{P}_{j,\text{ben}}$ and $\mathbf{P}_{j,\text{mal}}$ (for all $j = 1, \dots, N_{\text{val}}$ SAXS images in the validation data set) which are separately averaged to produce the naive triplet of probabilities $[\hat{\text{Prob}}_{j,\text{norm}}, \hat{\text{Prob}}_{j,\text{ben}}, \hat{\text{Prob}}_{j,\text{mal}}]$.

Section 8.3.3 describes how all N_{val} naive probabilities for each of the three groups is used to construct the raw histograms $[h_n(\hat{\text{Prob}}_{\text{norm}}), h_b(\hat{\text{Prob}}_{\text{ben}}), h_m(\hat{\text{Prob}}_{\text{mal}})]$ and Section 8.3.4 details how to smooth these histograms using a Walsh wavelet packet smoother in order to estimate the probability density functions $[\hat{p}(\hat{\text{Prob}}_{\text{norm}}), \hat{p}(\hat{\text{Prob}}_{\text{ben}}), \hat{p}(\hat{\text{Prob}}_{\text{mal}})]$.

Section 8.3.5 discusses how these probability density function estimates $[\hat{p}(\hat{\text{Prob}}_n), \hat{p}(\hat{\text{Prob}}_b), \hat{p}(\hat{\text{Prob}}_m)]$ are refined using an *adjustment* SAXS image data set. Finally, Section 8.3.6 provides an overall summary of the methodology.

Recall that parametric or non-parametric probability density functions $p_{g_{\text{norm}}}(z_{l,a,x,y})$, $p_{g_{\text{ben}}}(z_{l,a,x,y})$, $p_{g_{\text{mal}}}(z_{l,a,x,y})$ were estimated in Section 8.2 using a training data set, which were then used in Bayes' rule to estimate the posterior probabilities $[\hat{\text{Prob}}_n(z_{l,a,x,y}^j), \hat{\text{Prob}}_b(z_{l,a,x,y}^j), \hat{\text{Prob}}_m(z_{l,a,x,y}^j)]$ of group membership for each adaptive image transform coefficient $z_{l,a,x,y}$ for all the samples in a validation data set. These probabilities were then stored in probability arrays $\mathbf{P}_{j,\text{norm}}$, $\mathbf{P}_{j,\text{ben}}$ and $\mathbf{P}_{j,\text{mal}}$. It is desirable to reduce these vast arrays of posterior probabilities $\mathbf{P}_{j,\text{norm}}$, $\mathbf{P}_{j,\text{ben}}$ and $\mathbf{P}_{j,\text{mal}}$ into a simple overall probability $[\hat{\text{Prob}}_{j,\text{norm}}, \hat{\text{Prob}}_{j,\text{ben}}, \hat{\text{Prob}}_{j,\text{mal}}]$ for the group (or tissue type) which had been determined *a priori*.

Several approaches to produce these overall probability estimates are possible. One could consider each array as a sampled function and apply the techniques of functional data analysis. Unfortunately, models of functional data of such high-dimension at present do not exist or at best in their early stages of development. A hierarchical or similar regression model could also be developed, but at least two challenges are encountered:

- (i) a model to describe the general structure of the ragged arrays $\mathbf{P}_{j,\text{norm}}$, $\mathbf{P}_{j,\text{ben}}$ and $\mathbf{P}_{j,\text{mal}}$ must be found.
- (ii) a vast number of regression parameters might need to be estimated with limited data. Even if a model could be specified, it might be very difficult to compute the parameters of the model and obtain sufficient samples in practice to estimate them with sufficient accuracy.

A third approach which was adopted in this thesis is to obtain naive probability estimates by averaging of the probability arrays $\mathbf{P}_{j,\text{norm}}$, $\mathbf{P}_{j,\text{ben}}$ and $\mathbf{P}_{j,\text{mal}}$ and then to calibrate these naive probabilities using samples with known group labels from an independent data set. On first inspection, the naive probability estimates seemed too simplistic to be useful but it was found in practice that they could be used to develop more refined probability estimates. The key idea was to develop a transformation function that maps these naive probability estimates to their “correct” values which take into account the inherent correlations in the data. This straightforward approach may well be used as a baseline in the future to compare the accuracy of more sophisticated models.

The technique is to produce probability density function estimates of the naive posterior probabilities of tissue group (calculated for each SAXS image in the validation data set). These probability density function estimates are to be optimal in classifying, by Bayes’ rule, the groups of another independent data set called the *adjustment* data set. When satisfied with these probability density function estimates, they are then used to classify test data. The probability density function estimates are calculated using the Walsh wavelet packet which is specifically designed for data on the $[0, 1]$ domain.

For each SAXS image a naive probability is calculated as the simple average of all the individual posterior probabilities from each filter, scale and location. We appreciate that this transgresses fundamental statistical principles (e.g. they are not mutually exclusive) but these simple averages will be ingredients in the initial probability density function estimates for the posterior probability of group membership.

$$\begin{aligned}
\hat{\text{Prob}}_{j,\text{norm}} &= \frac{1}{Z} \sum_{l \in L} \sum_{a \in A} \sum_{x \in X} \sum_{y \in Y} \mathbf{P}_{j,\text{norm}}[l, a, x, y], \\
\hat{\text{Prob}}_{j,\text{ben}} &= \frac{1}{Z} \sum_{l \in L} \sum_{a \in A} \sum_{x \in X} \sum_{y \in Y} \mathbf{P}_{j,\text{ben}}[l, a, x, y], \\
\hat{\text{Prob}}_{j,\text{mal}} &= \frac{1}{Z} \sum_{l \in L} \sum_{a \in A} \sum_{x \in X} \sum_{y \in Y} \mathbf{P}_{j,\text{mal}}[l, a, x, y]
\end{aligned} \tag{8.16}$$

where L denotes the set of filter function indices, A denotes the set of scale indices, X the set of x-coordinates, Y the set of y-coordinates and $Z = [n(L) \times n(A) \times n(X) \times n(Y)]$ is the total number of coefficient entries or equivalently the product of the cardinality (denoted $n(\cdot)$) of the sets L , A , X , and Y that were retained in the adaptive image transform post application of the filter function selection algorithm in Section (7.4).

This naive triplet $[\hat{\text{Prob}}_{j,\text{norm}}, \hat{\text{Prob}}_{j,\text{ben}}, \hat{\text{Prob}}_{j,\text{mal}}]$ of overall posterior probabilities of each group for each of the $j = 1, \dots, N_{\text{val}}$ images is adjusted to improve classification performance by calculating yet another set of three univariate probability density functions $[p(\hat{\text{Prob}}_{\text{n}}), p(\hat{\text{Prob}}_{\text{b}}), p(\hat{\text{Prob}}_{\text{m}})]$ that encode the distribution of the naive overall posterior probability estimates for each group. Each of the three univariate probability density functions $[p(\hat{\text{Prob}}_{\text{n}}), p(\hat{\text{Prob}}_{\text{b}}), p(\hat{\text{Prob}}_{\text{m}})]$ is estimated using a Walsh wavelet packet smoother.

The Walsh wavelet packet smoother is discussed in Sections 8.3.4 and 8.3.5. It is specifically designed to estimate probability density functions from one-dimensional histogram data that is defined only on the domain $[0, 1]$. Each of the three univariate probability density functions $[p(\hat{\text{Prob}}_{\text{n}}), p(\hat{\text{Prob}}_{\text{b}}), p(\hat{\text{Prob}}_{\text{m}})]$ is estimated by applying the Walsh wavelet packet smoother to three separate one-dimensional histograms.

These one-dimensional histograms are created by considering the naive probabilities for each group separately. For instance, for the $j^* = 1, \dots, N_{\text{val}}$ images in the validation set we create the univariate histogram $h_n(\hat{\text{Prob}}_n)$ by considering only the naive probabilities $\hat{\text{Prob}}_{j^*,n}$. The univariate histogram $h_n(\hat{\text{Prob}}_n)$ is created by assigning the number of times that $\hat{\text{Prob}}_{j^*,n}$ occurs in the validation data set to the position $\text{Prob}_n = \text{Prob}_{j^*,n}$. A similar process is used to create the $h_b(\hat{\text{Prob}}_b)$ and $h_m(\hat{\text{Prob}}_m)$ histograms.

The domain of the histograms $[h_n(\hat{\text{Prob}}_n), h_b(\hat{\text{Prob}}_b), h_m(\hat{\text{Prob}}_m)]$ is the interval $([0, 1])$ which is then partitioned into 2^M sub-intervals (for instance when $M = 6$, 64 sub-intervals are created). A value of zero is initially assigned to each of these sub-intervals. A value of one is then added for each observation (either $\hat{\text{Prob}}_n$, $\hat{\text{Prob}}_b$ or Prob_m) that falls within a particular sub-interval. The Walsh wavelet packet transform was then applied to each of these pre-processed raw histograms.

Section 8.3.5 describes how the Walsh wavelet packet transform coefficients are ordered in terms of decreasing energy (squares of coefficients) and how only a proportion (say 25%) of the largest energy coefficients and the rest set to zero. The section goes on to explain how back-transformation yields smooth probability density function estimates and how to select the proportion of coefficients to retain (in each group) in order to achieve the most accurate classification on an independent adjustment SAXS image data set.

Once the probability density functions $[p(\hat{\text{Prob}}_n), p(\hat{\text{Prob}}_b), p(\hat{\text{Prob}}_m)]$ have been estimated, they can be used in conjunction with Bayes' rule (equation 8.2) to estimate the overall probability of group membership using the naive probabilities $[\text{Prob}_{j,\text{norm}}, \hat{\text{Prob}}_{j,\text{ben}}, \hat{\text{Prob}}_{j,\text{mal}}]$.

8.3.2 Creating the Naive Probability Matrix

The validation data set is used to create the probability arrays $\mathbf{P}_{\text{norm},j}$, $\mathbf{P}_{\text{ben},j}$ and $\mathbf{P}_{\text{mal},j}$ where each entry $\mathbf{P}_{\text{norm},j}(l, a, x, y)$, $\mathbf{P}_{\text{ben},j}(l, a, x, y)$ and $\mathbf{P}_{\text{mal},j}(l, a, x, y)$ corresponds to the probability estimates $\hat{\text{Prob}}_{\text{norm}}(z_{l,a,x,y}^j)$, $\hat{\text{Prob}}_{\text{ben}}(z_{l,a,x,y}^j)$, and $\hat{\text{Prob}}_{\text{mal}}(z_{l,a,x,y}^j)$ of the normal, benign and malignant tissue groups respectively. Each array $\mathbf{P}_{\text{norm},j}$, $\mathbf{P}_{\text{ben},j}$, $\mathbf{P}_{\text{mal},j}$ encodes the probability estimates for each three groups for all filters, scales and locations that were retained in the adaptive image transform post filter function selection. Three such arrays are produced for each SAXS image in the validation data set, the end result is a potentially large data set of arrays of probability estimates. This probability data set has the form,

$$\mathbf{P}_{\text{val}} = \begin{bmatrix} \mathbf{P}_{\text{norm},1} & \mathbf{P}_{\text{ben},1} & \mathbf{P}_{\text{mal},1} \\ \mathbf{P}_{\text{norm},2} & \mathbf{P}_{\text{ben},2} & \mathbf{P}_{\text{mal},2} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_{\text{norm},N_{\text{val}}} & \mathbf{P}_{\text{ben},N_{\text{val}}} & \mathbf{P}_{\text{mal},N_{\text{val}}} \end{bmatrix}. \quad (8.17)$$

There is a need to reduce the size of the \mathbf{P}_{val} array and provide an overall probability estimate for each of the three groups given the j th SAXS image. Recall that equation (8.16) specified how to obtain the naive probability estimates $[\hat{\text{Prob}}_{j,\text{norm}}, \hat{\text{Prob}}_{j,\text{ben}}, \hat{\text{Prob}}_{j,\text{mal}}]$ by simple averaging. The naive probabilities are stored in a N_{val} rows by three columns matrix, \mathbf{NP}_{val} of the form,

$$\mathbf{NP}_{\text{val}} = \begin{bmatrix} \hat{\text{Prob}}_{1,\text{norm}} & \hat{\text{Prob}}_{1,\text{ben}} & \hat{\text{Prob}}_{1,\text{mal}} \\ \hat{\text{Prob}}_{2,\text{norm}} & \hat{\text{Prob}}_{2,\text{ben}} & \hat{\text{Prob}}_{2,\text{mal}} \\ \vdots & \ddots & \vdots \\ \hat{\text{Prob}}_{N_{\text{val}},\text{norm}} & \hat{\text{Prob}}_{N_{\text{val}},\text{ben}} & \hat{\text{Prob}}_{N_{\text{val}},\text{mal}} \end{bmatrix}. \quad (8.18)$$

The entries (naive probability estimates) inside the \mathbf{NP}_{val} matrix will be used to estimate another set of probability density functions that are capable of an overall classification of group on future test data.

8.3.3 Creating the Raw Histograms

The \mathbf{NP}_{val} matrix is used to construct three raw histograms $h_n(\hat{\text{Prob}}_{\text{norm}})$, $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$. These raw histograms will be smoothed to produce the probability density functions estimates $\hat{p}(\hat{\text{Prob}}_{\text{norm}})$, $\hat{p}(\hat{\text{Prob}}_{\text{ben}})$ and $\hat{p}(\hat{\text{Prob}}_{\text{mal}})$. The raw histogram $h_n(\hat{\text{Prob}}_{\text{norm}})$ is constructed using all of the $\hat{\text{Prob}}_{j,\text{norm}}$ ($j = 1, \dots, N_{\text{val}}$) entries along the first column of the \mathbf{NP}_{val} matrix. Similarly, the raw histograms $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$ are constructed using the entries along the second and third columns of the \mathbf{NP}_{val} matrix respectively.

Each of the raw histograms $h_n(\hat{\text{Prob}}_{\text{norm}})$, $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$ are constructed by partitioning the interval $[0, 1]$ into 2^M segments. The power M is selected such that the number 2^M is the closest power of two possible to the number of samples N_{val} for the group g_{n^*} whose histogram is being estimated.

For instance, if the normal group in the validation data set had 60 samples then the power M would be selected to be six, as $2^6 = 64$. The interval $[0, 1]$ would then be partitioned into 64 segments and a value of zero assigned to each of these segments. For the raw $h_n(\hat{\text{Prob}}_{\text{norm}})$ histogram a value of ‘1’ is assigned to the segment that contains the $\mathbf{NP}_{\text{val}}(1, 1)$ magnitude. This process is repeated for all of the remaining $j = (2, \dots, N_{\text{val}})$, $\mathbf{NP}_{\text{val}}(j, 1)$ entries with a magnitude of ‘1’ added for each segment that the entry falls. The other raw histograms $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$ are constructed in a similar manner. The raw histograms $h_n(\hat{\text{Prob}}_{\text{norm}})$, $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$ are now ready for smoothing.

8.3.4 Transformation Of The Raw Histograms Using Walsh Wavelet Packets

The probability density function estimates $\hat{p}(\hat{\text{Prob}}_{\text{norm}})$, $\hat{p}(\hat{\text{Prob}}_{\text{ben}})$ and $\hat{p}(\hat{\text{Prob}}_{\text{mal}})$ are found by smoothing the raw histograms $h_n(\hat{\text{Prob}}_{\text{norm}})$, $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$. Smoothing is achieved by modification of the magnitude of coefficients from the wavelet packet transforms of the raw histograms, followed by back transformation. Because the raw histograms $h_n(\hat{\text{Prob}}_{\text{norm}})$, $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$ have a domain on the interval $[0, 1]$, there is a need to use specific smoothing functions.

The Walsh wavelet packet, which is a wavelet packet transform that uses Walsh basis functions, is based upon a family of analysis/smoothing functions that do not extend outside the interval $[0, 1]$. The Walsh wavelet packet transform also allows for adaptive smoothing of raw histograms. The Walsh basis is a family of functions derived from linear combinations of the Haar scaling function $\phi(x) = 1$ ($0 \leq x \leq 1$), 0 otherwise and the Haar wavelet function $\psi(x) = 1$ ($0 \leq x < 1/2$), -1 ($1/2 < x < 1$), 0 otherwise (page 133, Vidakovic 1999). The initial level of the Walsh basis functions $W_0(x) = \phi(x)$ and $W_1(x) = \psi(x)$ are simply the Haar scaling and wavelet functions respectively. The Walsh basis functions, $W_{2n}(x)$ and $W_{2n+1}(x)$, at higher levels ($n \geq 1$) are defined recursively as,

$$\begin{aligned} W_{2n}(x) &= W_n(2x) + W_n(2x - 1) \\ W_{2n+1}(x) &= W_n(2x) - W_n(2x - 1), \quad n \geq 1 \end{aligned} \tag{8.19}$$

(page 133, Vidakovic 1999).

The index, n , in the Walsh basis function $W_n(x)$ indicates the number of zeroes in the function. The greater the number of zeroes, the finer the detail the function $W_n(x)$ can detect.

The wavelet packet transform was discussed in Section (5.2.4). Recall that it uses a library of functions of the form,

$$\mathcal{W}_{j,n,k}(x) = 2^{\frac{j}{2}} \mathcal{W}_n(2^j x - k) \tag{8.20}$$

where parameters $j \in \mathbb{Z}$, $n \in \mathbb{N}$ and $k \in \mathbb{Z}$ are used to represent a function $f(x) \in L_2(\mathbb{R})$ (equation 5.13, page 135, Vidakovic 1999). The parameter, $j \in \mathbb{Z}$ is the scale parameter, parameter $n \in \mathbb{N}$ is the sequence parameter and parameter $k \in \mathbb{Z}$ is the translation parameter. The functions $\mathcal{W}_{j,n,k}(x)$ must be an orthogonal basis of the space $L_2(\mathbb{R})$.

When a subset \mathbb{P} of the scale $j \in \mathbb{Z}$ and sequence parameters $n \in \mathbb{N}$ are selected such the *dyadic* intervals (intervals with observations that can be indexed by parameters $(j, n) \in \mathbb{P}$ composed from the sets $j \in \mathbb{Z}$ and $n \in \mathbb{N}$), $I_{j,n} = \{[2^j n, 2^j(n+1)], (j, n) \in \mathbb{P}\}$ form a *disjoint* and *countable* covering of the interval $[0, \infty)$ (in this case $[0, 1]$), then the function $f(x) \in L_2(\mathbb{R})$ can be represented as,

$$f(x) = \sum_{(j,n) \in \mathbb{P}} \sum_{k \in \mathbb{Z}} w_{j,n,k} \mathcal{W}_{j,n,k}(x) \quad (8.21)$$

(adapted from Theorem 5.3.3, page 138, Vidakovic 1999).

Now the Walsh wavelet packet transformation specifically uses the Walsh basis functions, $W_n(x)$ in place of general orthogonal basis functions $\mathcal{W}_n(x)$ in the wavelet packet transform of equations (8.20) and (8.21). The raw histograms $h_n(\hat{\text{Prob}}_{\text{norm}})$, $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$ are in practice represented by the discrete sequence of numbers (h_n) , (h_b) and (h_m) . These sequences are within the space of square summable sequences $l_2(\mathbb{R})$ as $\sum_{n=1}^{M_{\text{norm}}} |h_n|^2 < \infty$, $\sum_{n=1}^{M_{\text{ben}}} |h_b|^2 < \infty$ and $\sum_{n=1}^{M_{\text{mal}}} |h_m|^2 < \infty$. The space $L_2(\mathbb{R})$ is the discrete counterpart to the space $L_2(\mathbb{R})$. Therefore, discrete representations of the Walsh basis functions $W_n(x)$ can be used to obtain discrete representations of the probability density functions $\hat{p}(\hat{\text{Prob}}_{\text{norm}})$, $\hat{p}(\hat{\text{Prob}}_{\text{ben}})$ and $\hat{p}(\hat{\text{Prob}}_{\text{mal}})$.

It is convenient to describe the Walsh wavelet packet transforms of the raw histograms in terms of *crystals*. Crystals are vectors, denoted $\mathbf{w}_{j,n} = (w_{j,n,0}, w_{j,n,1}, \dots, w_{j,n,2^j-1})$ that contain all coefficients $\mathbf{w}_{j,n,k} = \langle f(x), W_{j,n,k}(x) \rangle$ where $f(x)$ is either $h_n(\hat{\text{Prob}}_{\text{norm}})$, $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$ whilst $W_{j,n,k}(x)$ is given by equation (8.20) when Walsh basis functions $W_n(2^j - k)$ are substituted for $\mathcal{W}_n(2^j x - k)$. A particular crystal $\mathbf{w}_{j,n}$ contains all the coefficients $w_{j,n,k}$ at a fixed scale j and sequence parameter, n . The individual entries $w_{j,n,k}$ are called *atoms* (page 138, Vidakovic 1999). The crystals from the Walsh wavelet packet transform can be depicted graphically as in Figure 8.9.

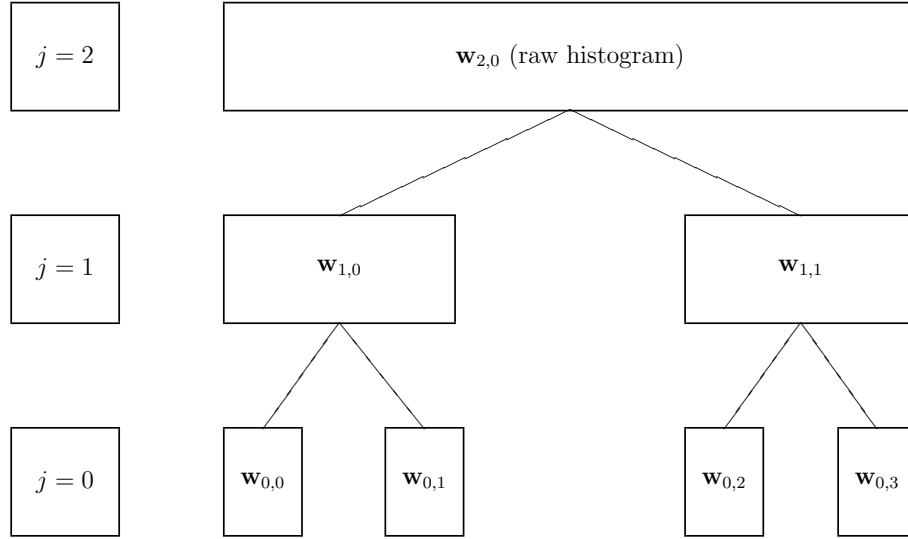


Figure 8.9: Crystal organisation in a two-level Walsh wavelet packet transform.

Level 2 of the Walsh wavelet packet transform is denoted by the crystal $w_{2,0}$ and corresponds to the raw histogram data (either $h_n(\hat{\text{Prob}}_{\text{norm}})$, $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$). The Walsh basis functions $w_{1,0,1}, \dots, w_{1,0,2^M}$ are applied to the crystal $w_{2,0}$ to yield the crystal $w_{1,0}$ after down-sampling. Similarly the Walsh basis functions $w_{1,1,1}, \dots, w_{1,1,2^M}$ are applied to the crystal $w_{2,11}$. The process continues throughout the tree depicted in Figure 8.9 with the Walsh basis functions $w_{j,n,1}, \dots, w_{j,n,2^M}$ are used to obtain the crystal $w_{j,n}$.

The Walsh wavelet packet transform applies the full wavelet packet decomposition using Walsh basis functions. In other words, all crystals are retained. In order to achieve adaptive smoothing on the interval only certain atoms will be retained. All other atoms will be set to zero. The Walsh wavelet packet transform is then back transformed to yield the smoothed estimate of the probability density function. Selection of which atoms to retain and which to set to zero is based upon classification performance of the smooth probability density function estimates on an adjustment data set. The methodology of this selection process will be the subject of the next section.

8.3.5 Smoothing of the Raw Histograms Using the Walsh Wavelet Packet Transform

The Walsh wavelet packet transform was used to analyse the raw histograms $h_n(\hat{\text{Prob}}_{\text{norm}})$, $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$ with a range of different basis functions. These basis functions are useful for extracting different features from the raw histograms, a feature is detected when a relatively large magnitude coefficient is produced. In order to smooth the raw histograms $h_n(\hat{\text{Prob}}_{\text{norm}})$, $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$ to produce the probability density function estimates $\hat{p}(\hat{\text{Prob}}_{\text{norm}})$, $\hat{p}(\hat{\text{Prob}}_{\text{ben}})$ and $\hat{p}(\hat{\text{Prob}}_{\text{mal}})$ the magnitude of the Walsh wavelet packet transform coefficients needs to be modified. Retaining only the largest magnitude coefficients allows the key features that the Walsh wavelet packet transform has detected in the raw histogram to be retained and for the remaining variation discarded when estimating the probability density functions.

The challenge for successful probability density function estimation is how to select which coefficients to retain and which to omit. A percentile thresholding scheme is used to select Walsh wavelet packet coefficients. The actual percentile used is determined in a step wise fashion.

First threshold the coefficients by selecting those whose energies are greater than the q quantile of the sample. At the start, q is set to 0.5, 0.625 and 0.75. The inverse transform of the retained Walsh coefficients gives three versions of probability density function estimates for each set of naive probabilities $\hat{\text{Prob}}_{\text{norm}}$, $\hat{\text{Prob}}_{\text{ben}}$, $\hat{\text{Prob}}_{\text{mal}}$.

These probability density function estimates are now used in Bayes' rule to classify tissue group in the adjustment data set and the three which perform best are kept and the others discarded. An example of this process is depicted in Figure 8.10. The combinations scoring '1' would be kept, indicating the appropriate level of thresholding which is denoted in Figure 8.10 as q^* , q^{**} and q^{***} .

		0.500	0.625	0.750
\hat{p}_{norm}	*	1	2	3
\hat{p}_{ben}	**	3	1	2
\hat{p}_{mal}	***	1	3	2

Figure 8.10: Determining the most accurate probability density function estimates using an adjustment data set. Three probability density function estimates are determined for each of the three groups by retaining select quantiles of Walsh wavelet packet transform coefficients followed by back-transformation. The classification performance is assessed on the adjustment data set using the different combinations of probability density function estimates and Bayes' rule. The probability density function estimates in each group that produce the most accurate classification on the adjustment data set (as indicated by '1') are retained and the other probability density function estimates discarded.

We return to the naive estimates of posterior probabilities and repeat the process with three new quantiles, centered around those retained in the previous step. The set of probability density function estimates that are selected in this refinement process are retained for classification of test data.

The explicit algorithm used in this thesis is as follows:

- i) Obtain the raw histograms $h_n(\hat{\text{Prob}}_{\text{norm}})$, $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$.
- ii) Apply the Walsh wavelet packet transform to each histogram.
- iii) In each case, order the Walsh wavelet packet transform coefficients from greatest to smallest energy.
- iv) Select the q th quantile of largest energy coefficients, where parameter q is obtained from the set $Q = [0.250, 0.375, 0.500]$.
- v) Set the remaining Walsh wavelet packet transform coefficients to zero.
- vi) Back transform the modified Walsh wavelet packet coefficients for each quantile and each group.
- vii) Compare the classification accuracy using Bayes' rule for all combinations of the probability density function estimates between groups.

- viii) Select the probability density function estimates that maximise $\text{Prop}_{q^*,n}$, $\text{Prop}_{q^{**},b}$ and $\text{Prop}_{q^{***},m}$.
- ix) Refine estimates $\hat{p}_{q^*}(\hat{\text{Prob}}_{\text{norm}})$, $\hat{p}_{q^{**}}(\hat{\text{Prob}}_{\text{ben}})$ and $\hat{p}_{q^{***}}(\hat{\text{Prob}}_{\text{mal}})$. Repeat as from step ii) but this time use the quantiles,

$$\begin{aligned}
 q_{\text{norm}} &= \{q^* - 0.0625, q^*, q^* + 0.0625\} \\
 q_{\text{ben}} &= \{q^{**} - 0.0625, q^{**}, q^{**} + 0.0625\} \\
 q_{\text{mal}} &= \{q^{***} - 0.0625, q^{***}, q^{***} + 0.0625\}.
 \end{aligned}
 \tag{8.22}$$

Use steps (iv) and (v) again.

A larger number of entries could have been included in the first set Q in order to consider at once a wide range of probability density function estimates. This could not be done in practice as the Walsh wavelet packet transform method of estimating probability density functions requires a large number of samples when a large number of probability density function estimates are to be compared simultaneously. If a greater number of probability density function estimates were compared simultaneously without a substantial increase in sample size then comparisons between the probability density function estimates would be meaningless. This problem is produced by the $\text{Prop}_{q^*,n}$, $\text{Prop}_{q^{**},b}$ and $\text{Prop}_{q^{***},m}$ statistics having a high variability when the sample size is insufficient. Therefore, the number of comparisons was constrained to limit the impact of this problem on the probability density function estimation process. In practice, it was found that the Walsh wavelet packet transform had such few large magnitude coefficients that the omission of smaller energies (below the 0.50th quantile) was of little relevance. Furthermore it was also found in practice that it was not necessary to further refine the probability density function estimates produced in step (ix) above .

8.3.6 Summary and Conclusion to Estimating the Overall Posterior Probability of each Group

This section outlined a method to produce three estimates of the overall probability that a tissue sample belongs to either the normal, benign or malignant breast tissue groups based upon the probability arrays produced in Section 8.2 using the coefficients of the adaptive image transform of its SAXS image. The central theme was to construct univariate histograms $h_n(\hat{\text{Prob}}_{\text{norm}})$, $h_b(\hat{\text{Prob}}_{\text{ben}})$ and $h_m(\hat{\text{Prob}}_{\text{mal}})$ and to apply the Walsh wavelet packet transform in each case. A proportion of the coefficients was retained and the remaining set to zero. Smoothed probability density function estimates were obtained following back transformation of the modified Walsh wavelet packet transform coefficients. A number of different probability density function estimates were trialled for each group and a subset of three (one for each group) was selected based upon their classification accuracy on an independent adjustment SAXS image data set. The model was then ready to be applied to actual test data.

8.4 Summary & Conclusion to Statistical Modeling

A range of new statistical techniques have been introduced in this chapter in order to calculate the scale-location and overall group probability estimates for the SAXS image data set. Notable contributions include the XOR model, the Mexican hat contourlet transform, the bivariate probability density function estimation using the generalised extreme value distribution function and similarly using the Walsh wavelet packet transform to perform adaptive univariate probability density function estimation. A few comments are in order for each of these techniques.

The XOR model is designed for large-scale image analysis applications, it was proposed as a potential solution when a very large number of probability density function estimates must be calculated and *stored* in the computer memory. Kernel density estimates or other density estimation procedures could have been applied but there are four reasons why we have used and advocated the XOR model approach:

- a) There is good reason to believe that many of the probability density functions (conditional on group) of the adaptive image transform coefficients can be well-modeled by a bivariate Gaussian distribution (at least for SAXS images). Storing the parameters of a bivariate Gaussian distribution rather than that of a bivariate kernel estimate is very likely to require far less computer memory. This saving in memory might result in a greater computational efficiency for routine large scale applications of the final diagnostic model.
- b) Any parametric probability density function can be used in the parametric component of the XOR model. All that needs to be done is to replace the Shapiro-Wilk test with another goodness of fit test that is suitable to assess the adequacy of the probability density function that is proposed.
- c) XOR model also allows non-parametric probability density function estimation using any method. In particular, bivariate probability density function estimates can be found using the Mexican hat contourlet transform. Raw histograms of adaptive image transform coefficient magnitudes are smoothed by modifying the Mexican hat contourlet transform coefficients using the statistical theory of extreme values. A large proportion of these smoothed coefficients are zero which allows for sparse probability density function representations.

- d) This thesis is concerned as much about the development of new multi-scale statistical image analysis as it is with the development of classification model of SAXS images. Probability estimates and function selection could have been achieved using standard statistical methods (such as with kernel density estimates) but this would have negated the opportunity to explore both the XOR density estimation model and non-parametric smoothing using the Mexican hat contourlet transform.

A disadvantage of the XOR model is that it often requires many applications of the Shapiro-Wilk test combined with a correction to avoid an unnecessarily high false-discovery rate. Statistical criterion that can decide between a parametric or non-parametric model thereby avoiding the need for a large number of tests would be of great advantage in this setting.

The Mexican hat contourlet transform was developed to in order to combine a wavelet filter with a directional filter. This allowed both the edges and their directions to be detected when analysing bivariate data. The Mexican hat contourlet transform has the advantage over other multi-scale and multi-directional bivariate data transforms such as the ordinary contourlet and curvelet transforms in that it requires fewer coefficients to describe the same data. This has the advantage of reducing the computational and storage requirements when analysing bivariate data. Such a property is advantageous when large numbers of bivariate data must be transformed and analysed, as in the case of the SAXS image analysis model presented in this chapter.

The Mexican hat contourlet transform allowed raw histogram data to be analysed using a range of functions that varied in scale, orientation (direction) and location. The different scale and direction matrices of the Mexican hat contourlet transform are similar to smoothing the original bivariate data with different smoothing functions. Large magnitude coefficients were only produced when there was a good match between the Mexican hat contourlet smoothing function and the raw histogram data.

Computer simulation studies on the absolute magnitude of coefficients from the Mexican hat contourlet transforms of both random data and probability density functions allowed coefficient smoothing models to be developed based upon the generalised extreme value distribution. Coefficient weights were determined from the generalised extreme value distributions. These weights were multiplied with the original Mexican hat contourlet transform coefficient magnitude for raw histograms produced using the adaptive image transform coefficients from the training SAXS image data set. The modified coefficients were then back transformed using the inverse Mexican hat contourlet transform to yield the smooth probability density function estimate.

Because computer simulations were used in the estimation of the generalised extreme value distributions, any type of model could have been proposed to produce the random data. That is, using the Mexican hat contourlet transform combined with the generalised extreme value distribution functions allows smoothing to be performed under any type of noise model. Further research into this technique would be useful to compare the bias and variance of the resulting probability density function estimates with those produced by other probability density function estimation techniques.

Finally, estimation of univariate probability density functions on the interval was achieved using the Walsh wavelet packet transform. The support of the Walsh basis functions was on the interval. This removed the problem of inaccurate probability density function estimates caused by using a smoothing function that extends outside the domain of definition. Whilst use of the wavelet packet transform allowed for an analysis of the raw histogram data with a range of different functions. Such an approach allowed different features to be detected in the raw histograms of naive probabilities. The use of a separate adjustment data set provided the opportunity to estimate probability density functions (by retaining only a subset of Walsh wavelet packet transform coefficients) that produced the best classification results.

Chapter 9

Application to Breast Cancer Diagnosis

This chapter concerns the application of the image analysis and statistical modeling methodology that was presented in Chapters 7 and 8 to SAXS image data. The objective of this chapter is to assess the utility of these techniques in the identification of breast cancer. Qualitative and quantitative results will be presented that allow for an appreciation of how the methods developed in this thesis work in practice.

Section 9.1 presents the results that pertain to the image analysis methodology that was developed in Chapter 7. The coefficients $z_{l,a,x,y}$ of the adaptive image transform of SAXS images are reviewed and coefficient matrices are compared for SAXS images that are associated with different tissue pathologies. The probability of misclassification is calculated for each filter function used in the adaptive image transform and used as a guide to assess classification model performance.

Section 9.2 involves the application of the diagnostic models that were developed in Section 8.2. The XOR model of Section 8.2.2 was implemented with bivariate Gaussian probability density functions being fitted to describe the adaptive image transform coefficient $z_{l,a,x,y}$ coefficient magnitude when appropriate and non-parametric probability density function estimates used when they were not. The Mexican hat contourlet transform of Section 8.2.4 was used to smooth the raw histograms $f_{\text{raw}} [\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ by modification of the transform coefficients using generalised extreme value distribution functions $[\hat{P}(z, \mu_1, \sigma_1, \xi_1), \hat{P}(z, \mu_2, \sigma_2, \xi_2), \dots, \hat{P}(z, \mu_J, \sigma_J, \xi_J)]$ as per the methodology of Section 8.2.8.4.

Section 9.3 concerns the implementation of the methodology developed in Section 8.3. The naive probabilities $[\hat{\text{Prob}}_{j,\text{norm}}, \hat{\text{Prob}}_{j,\text{ben}}, \hat{\text{Prob}}_{j,\text{mal}}]$ were calculated and the raw histograms $[h_n(\hat{\text{Prob}}_{\text{norm}}), h_b(\hat{\text{Prob}}_{\text{ben}}), h_m(\hat{\text{Prob}}_{\text{mal}})]$ constructed using the methods of Sections 8.3.2 and 8.3.3. The coefficients of the Walsh wavelet packet transforms of these histograms was then modified using the thresholding procedure discussed in Section 8.3.4. The inverse Walsh wavelet packet transform was then applied to yield the probability density function estimates $[\hat{p}(\hat{\text{Prob}}_n), \hat{p}(\hat{\text{Prob}}_b), \hat{p}(\hat{\text{Prob}}_m)]$. Several probability density function estimates were calculated and compared by retaining different numbers of Walsh wavelet packet coefficients in each case. The three probability density function estimates that produced the best classification performance on a separate adjustment data set were retained as per Section 8.3.5.

Section 9.4 applied the bivariate and univariate probability density function estimates developed in the previous two sections as a complete model on an independent test data set. This allowed for an objective assessment of the diagnostic model in practice. Finally, Section 9.5 ends the chapter with a summary and conclusion.

9.1 Application of the Adaptive Image Transform to SAXS Images

9.1.1 SAXS Image Data Set

	N	B	M
Training	50	50	50
Validation	50	49	50
Adjustment	50	50	50
Test	10	10	10

Table 9.1: Number of tissue samples in each of the normal (N), benign (B) and malignant (M) tissue groups in each of the data sets used in model development and assessment.

As indicated in Chapter 8 model development requires that we partition the SAXS image data set in four subsets, the *training*, *validation*, *adjustment* and *test* data sets. The training data set is used to estimate the parametric or non-parametric bivariate probability density function estimates of the real and imaginary components of the adaptive image transform coefficient $z_{l,a,x,y}$ magnitude. The validation data set is used to calculate the posterior probabilities of the tissue sample belonging to the normal, benign or malignant tissue groups at each filter, scale and location of the adaptive image transform. This data set is also used to calculate the naive probabilities $[\hat{\text{Prob}}_{j,\text{norm}}, \hat{\text{Prob}}_{j,\text{ben}}, \hat{\text{Prob}}_{j,\text{mal}}]$, the raw histograms $[h_n(\hat{\text{Prob}}_{\text{norm}}), h_b(\hat{\text{Prob}}_{\text{ben}}), h_m(\hat{\text{Prob}}_{\text{mal}})]$ and the probability density function estimates $[\hat{p}(\hat{\text{Prob}}_n), \hat{p}(\hat{\text{Prob}}_b), \hat{p}(\hat{\text{Prob}}_m)]$ using the Walsh wavelet packet transform. The adjustment data set is used to select the best performing set of univariate probability density function estimates. This is achieved by trialling different combinations of probability density function estimates and using them to predict the group label of observations (within the adjustment data set). The predicted group labels are compared with the known group labels and the most accurate combination of univariate probability density functions estimates are retained. The test data set is an independent set of SAXS images held apart from the model development and is used to assess the entire model (consisting of the adaptive image transform as well as the bivariate and univariate probability density function estimates). A numerical summary of the number of normal (N), benign (B) and malignant (M) breast tissue samples in each data set is provided in Table 9.2.

9.1.2 Adaptive Image Transformation of SAXS Images within the Training Data Set

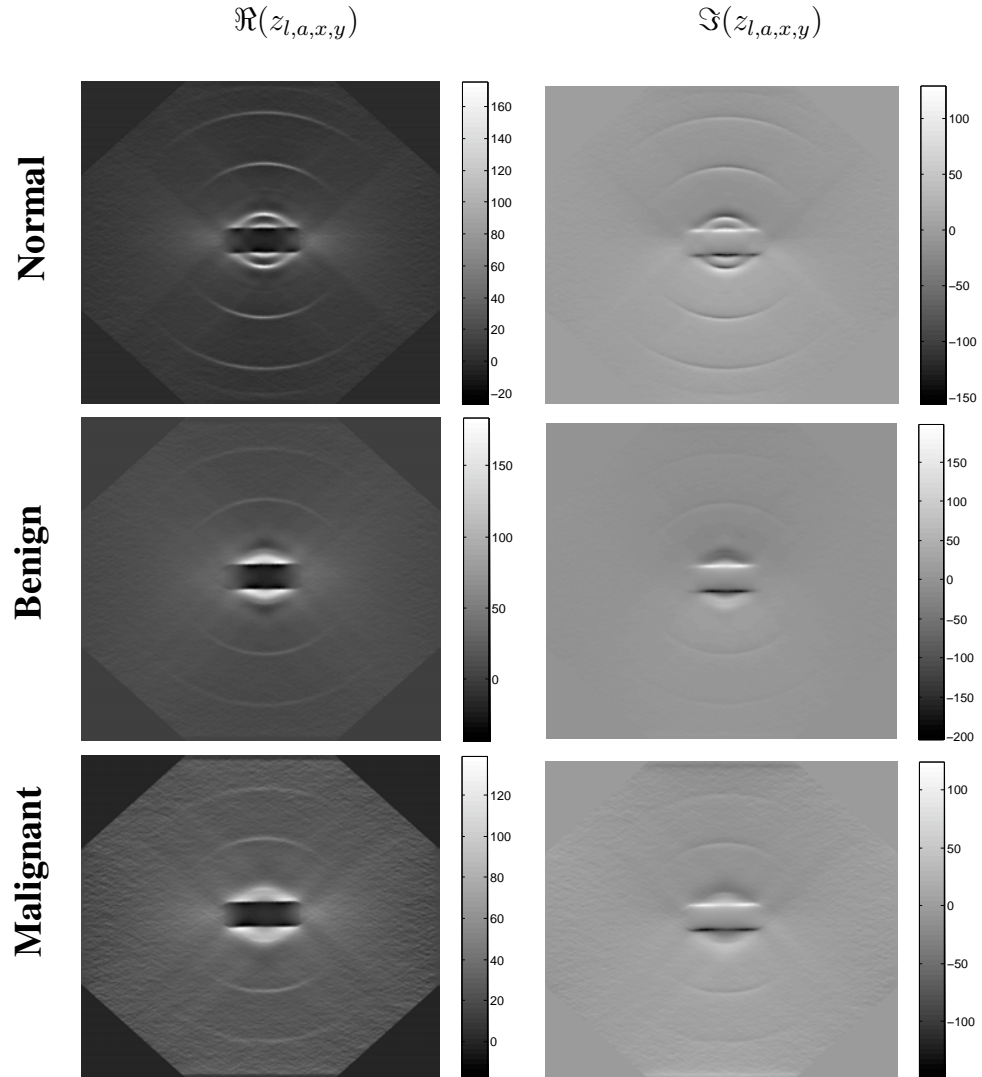
The adaptive transform of Section 7.2 was applied to the SAXS image training data set. The filters in the library of the transform included the Mexican hat, Pet hat, Gabor, Gamma, Mellin, Chebyshev, Zeta and Witch's Hat functions. The adaptive image transform was applied to each individual SAXS image using each function in the library at a range of scales ($a = 1, 2, 3, \dots, 8$).

Visible differences in the real and imaginary components of the coefficients $z_{l,a,x,y}$ were evident between SAXS images from different tissue groups. In particular, the magnitude of scattering ring features appeared different between the tissue groups. Figure 9.1 displays the real $\Re(z_{l,j,a,x,y})$ and imaginary $\Im(z_{l,j,a,x,y})$ components of the coefficients of the adaptive image transform using the Witch's Hat filter function at scale $a = 2$ for select SAXS images of normal, benign and malignant breast tissue.

The main difference between the coefficient matrices appears to be the magnitude of those coefficients that are associated with the meridional scattering rings. The set of coefficients associated with the normal breast tissue sample has a sharply defined set of rings in both the real ($\Re(z_{l,a,x,y})$) and imaginary ($\Im(z_{l,a,x,y})$) components. In contrast, the scattering ring features in the adaptive image transform coefficients from the SAXS image belonging to the benign sample do not appear as sharply defined. Note the apparent differences in the magnitude of the scattering rings and that the benign sample has a poorly defined 1st order meridional scattering ring which is located in the centre of the coefficient matrix. These differences appear greatest when comparing the imaginary coefficient matrices $\Im(z_{l,a,x,y})$ of the normal and benign samples.

Both components of the real and imaginary coefficients have distinct differences between the normal and malignant sample. The meridional scattering rings in the real coefficient component ($\Re(z_{l,a,x,y})$) of the malignant sample appear to have a smaller magnitude than those in the normal sample. The 1st order meridional scattering ring is also less well defined in the malignant sample and there appears to be a greater amount of variation (appears as a rougher surface) in coefficient magnitude in the regions of the coefficient matrix that are associated with amorphous scatter. The imaginary component ($\Im(z_{l,a,x,y})$) of the malignant sample also supports these observations with the 1st meridional order scattering ring again being poorly detected in contrast to that of the normal sample.

Figure 9.1: Witch's Hat filter function - adaptive image transform coefficients at scale ($a = 2$) for select SAXS images of normal, benign and malignant breast tissue.



The real and imaginary components of the benign and malignant tissue samples also appear to have distinct differences. The real component $\Re(z_{l,a,x,y})$ of the coefficients contains meridional scattering ring features that appear to have a greater magnitude than those in the real component of the coefficients of the malignant sample. A 1st order meridional scattering ring is visible in the coefficients of the malignant sample but in the coefficients of the benign sample this feature is very diffuse and barely visible. The real component of the coefficients of the malignant sample also has greater variation than the real component of the coefficients of the benign sample in the regions associated with amorphous scatter. The imaginary components $\Im(z_{l,a,x,y})$ of the coefficients support the above observations but also have more pronounced differences in magnitude. This apparent difference may act as an important discriminant between benign and malignant samples.

The limitations of this qualitative inspection of the coefficient matrices must be kept in mind. Only one example from each of the three tissue groups has been examined, furthermore, they have been examined for only one scale $a = 2$ and for one filter function (the Witch's Hat). The apparent differences presented could have arose from random chance or have been deliberately selected to present the case that the coefficients of the adaptive image transform pick out differences in breast tissue groups related to the meridional scattering ring and amorphous scatter features. To counter these difficulties the entire data set - all images, for all filters and scales were examined using the same pixel palette scale. The differences presented in Figure 9.1 are indicative of the general trend observed, sometimes the differences in the coefficient components was clear-cut (and related to the scattering ring features) and in some cases it was not. In all the inspection suggested that differences in the magnitude of the components of the adaptive image transform coefficients might be able to distinguish between SAXS images produced by different breast tissue pathologies. At this stage, there is no definitive evidence that the key differences are related to the meridional scattering ring and amorphous features, but in some instances visual inspection did suggest that this was the case.

Visual assessment and comparison of coefficient matrices is clearly too subjective and potentially misleading given the large number of comparisons that need to be made. Furthermore, subtle feature not readily distinguished by eye may be a key discriminant and complex interactions that are not readily deciphered might need to be accounted for when determining tissue group. The possibility that the adaptive image transform coefficients of the SAXS images cannot distinguish between breast tissue groups cannot yet be discounted. It is clear that statistical methodology (such as that developed in Chapter 8) needs to be implemented to obtain an objective assessment of this complicated data set.

Figure 9.2: Bayes' probability of misclassification across locations for the SAXS image dataset.

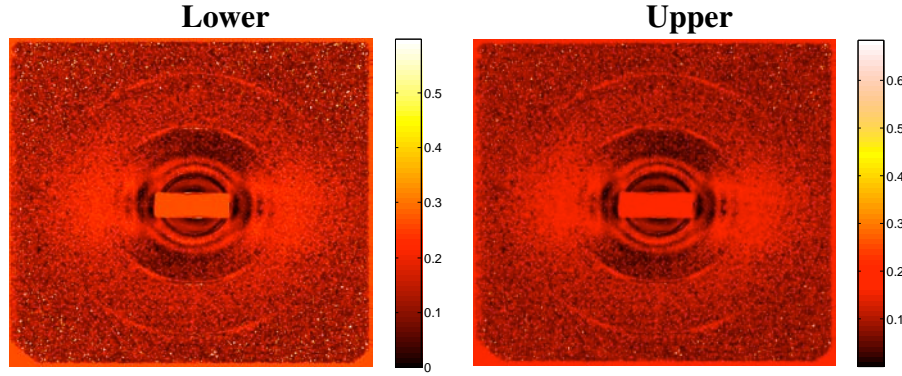


Figure 9.2 displays the upper and lower bounds on the Bayes' probability of misclassification across all locations at scale $a = 2$ of the adaptive image transform for the SAXS image training data set. In other words, the probability of misclassification displayed in Figure 9.2 is that of the combination of filter functions retained post the filter function selection algorithm in Section 7.4. Figure 9.2 displays the qualitative trends in the probability of misclassification (as calculated from Section 7.4) across locations for scale $a = 2$. Clear differences exist in the probability of misclassification varies across different locations in the SAXS images. Therefore selection of distinct locations of coefficients as *features* carries the risk of selecting regions that are inherently prone to greater classification error. Note that the equatorial scattering rings and the diffuse scatter region between the first and third meridional scattering rings has the lowest probability of error, suggesting that these regions although not easily identifiable by visual inspection of the coefficient matrices are the regions of the SAXS image most likely to produce accurate classification. Also observe that the 3rd and 5th meridional scattering ring region have considerably greater probability of error. This is of practical importance as the majority of SAXS image classification algorithms including those of Falzon *et al* (2006) have focused upon the extraction of coefficient features from the 3rd and 5th meridional scattering peaks. The region of the SAXS images that is potentially prone to the greatest probability of misclassification.

To further understand the effect of filter function on classification performance, the bounds on the probability of misclassification were examined a single filter function at a time rather than as a combination of filter functions as done in Figure 9.2. Table 9.2 reports on the range of the bounds of Bayes' probability of misclassification across *all* locations for the scales ($a = 2, 4$ and 6) and all of the filter functions in the adaptive image transform library using the SAXS image training data set. Reporting the bounds on the probability of misclassification for all filters, scales and locations would be too confusing and distracting. Therefore, Table 9.2 is understood to be a guide to the classification accuracy possible using the adaptive image transform coefficients of different filter functions at different scales.

Classification accuracy appears to be dependent upon both the filter function used for the analysis as well as the scale used in the transform. The Mexican hat wavelet and Witch's Hat filter functions appear to be very good choices for scale $a = 2$, both have a lower bound of 0.05 and upper bounds of 0.21 and 0.22 respectively. The results at the medium scale $a = 4$ suggests using the Gabor, Gamma and Mexican Hat filters, whilst the Gabor functions appear superior for scale $a = 6$ with error bounds of [0.05,0.21].

	Scale		
	$a = 2$	$a = 4$	$a = 6$
Gabor	[0.06,0.23]	[0.06,0.23]	[0.05,0.21]
Gamma	[0.18, 0.72]	[0.06,0.25]	[0.18,0.74]
Mellin	[0.18,0.72]	[0.18,0.73]	[0.18,0.73]
Mexican Hat	[0.05,0.21]	[0.06,0.26]	[0.15,0.60]
Pet Hat	[0.17,0.67]	[0.18,0.72]	[0.18,0.73]
Chebyshev	[0.18,0.73]	[0.16,0.64]	[0.69,0.99]
Witch's Hat	[0.05,0.22]	[0.18,0.72]	[0.18,0.72]
Zeta	[0.04,0.14]	[0.14,0.57]	[0.18,0.72]

Table 9.2: Bounds on probability of misclassification of the three tissue groups across scales ($a = 2, a = 4$ and $a = 8$) for the different filter functions in the adaptive image transform library.

Analysis of Table 9.2 supports the case that an adaptive image transform that uses a variety of filter functions and analyses the SAXS image over multiple scales is a valuable method. For instance, the Chebyshev filter function at scale $a = 6$, suggests poor classification performance should be expected as the bounds on probability of misclassification are $[0.69, 0.99]$. In contrast, at the same scale the Gabor filter has a probability of misclassification of $[0.05, 0.21]$. Selecting the Chebyshev filter function at this scale exclusively for the analysis of the SAXS images would have lead to poor classifier. Choice of the appropriate filter function appears to be very important. Note that filter function performance is also scale dependent. For instance the Zeta filter at scale $a = 2$ has bound on the probability of misclassification of $[0.04, 0.14]$, whilst at scale $a = 6$ this bound is $[0.18, 0.72]$.

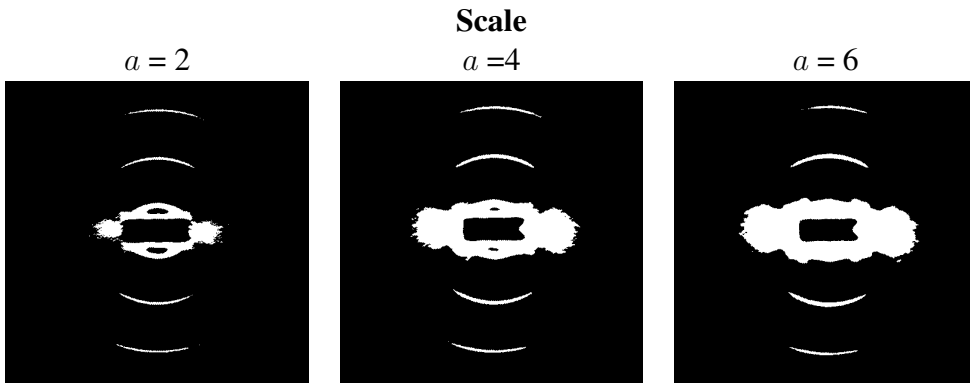
Based upon these results data dependent selection of the filter functions to use in the multi-scale analysis of SAXS images appears to be a wise strategy. By allowing a range of filter functions with different characteristics into the library of the adaptive image transform we might be able to capture a range of different, potentially useful diagnostic features. In the case of SAXS images of breast tissue, correct filter selection for each scale and location of the adaptive image transform can substantially alter the accuracy of the final model. Based upon the results presented in Table 9.2 the most accurate classification model that can be expected from the adaptive image transform coefficients is 96 %.

9.2 SAXS Breast Tissue Classification Model

This section concerns the implementation of the statistical methodology developed in Section 8.2. The XOR model is implemented and probability density functions of the adaptive image transform coefficient magnitudes estimated parametrically and non-parametrically using the Mexican hat contourlet transform. This is the first step in predicting tissue group from the adaptive image transform coefficients of a SAXS image. The outcome at the end of this section is the estimation of the probability of each tissue group for each filter, scale and location of the transform.

9.2.1 Implementation of the XOR model

Figure 9.3: The adaptive image transform coefficients (for the Witch's Hat filter and the normal group) across scales that were selected for non-parametric probability density function estimation using the XOR model (displayed in white).



The first step in the analysis of the adaptive image transform coefficients of the SAXS image training data set is to apply the XOR model of Section 8.2.2 in order to determine which bivariate probability density functions of coefficient magnitude need to be determined non-parametrically. Shapiro-Wilks tests were applied to all the adaptive image transform coefficients in each tissue group for each filter, scale and location. Those coefficient filter, scale and locations that were not well modelled using a bivariate Gaussian probability density function were selected for modelling with a non-parametric probability density function estimator. The Mexican hat contourlet transform was selected for performing this non-parametric probability density function estimation.

Figure 9.3 displays examples of the adaptive image transform coefficients selected for the normal tissue group and the Witch’s Hat filter function across all locations at select scales ($a = 2, 4, 6$). Coefficients that were selected for non-parametric probability density function estimation are displayed in white, those that were not selected either because they were not retained post filter-function selection (in Section 7.4) or because they were well modelled using a bivariate Gaussian probability density function are displayed in black. Inspection of Figure 9.3 reveals that those coefficient locations (for at least the Witch’s Hat filter and the scales $a = 2, 4, 6$) correspond to the same locations as the meridional and equatorial scattering ring features. This result matches well with the case presented in Section 8.2.3 in which it was stated that coefficients associated with amorphous features are well modelled with bivariate Gaussian probability density functions but those associated with scattering ring features are not. Also observe in Figure 9.3 that as the scale increases a greater ring of the equatorial scattering ring features appear to be captured. Such a result supports the hypothesis that scattering features exist in the SAXS images over a variety of scales and advocates the utility of multi-scale methods in the extraction of useful information in such cases.

Scale	Proportion
1	0.04
2	0.03
3	0.03
4	0.07
5	0.10
6	0.14
7	0.14
8	0.16

Table 9.3: Proportion of adaptive image transform coefficients (across groups, filters and locations) selected to be modelled using non-parametric probability density function estimates from scales $a = 1, \dots, 8$. All proportions are reported to two decimal places.

Table 9.2.1 reports on the total proportion of all adaptive image transform coefficients across groups, filter, scales and locations that were selected for non-parametric probability density function estimation. It is apparent that a large proportion of coefficients can be adequately modelled using the bivariate Gaussian probability density function. As the scale parameter increases the greater this proportion becomes which could be a result of a greater presence of equatorial scattering ring features at higher scales, an increase in the support and hence the region that the filter function analyses as scale increases or a combination of both effects. The total proportion of adaptive image transform coefficient indices across all groups, filters, scales and locations selected for modelling using non-parametric probability density function estimation was 0.09 (to two decimal places). Implementation of the XOR model greatly reduced the storage requirements of the SAXS image classification model, therefore achieving the objective for which it was designed.

9.2.2 Bivariate Probability Density Function Estimation

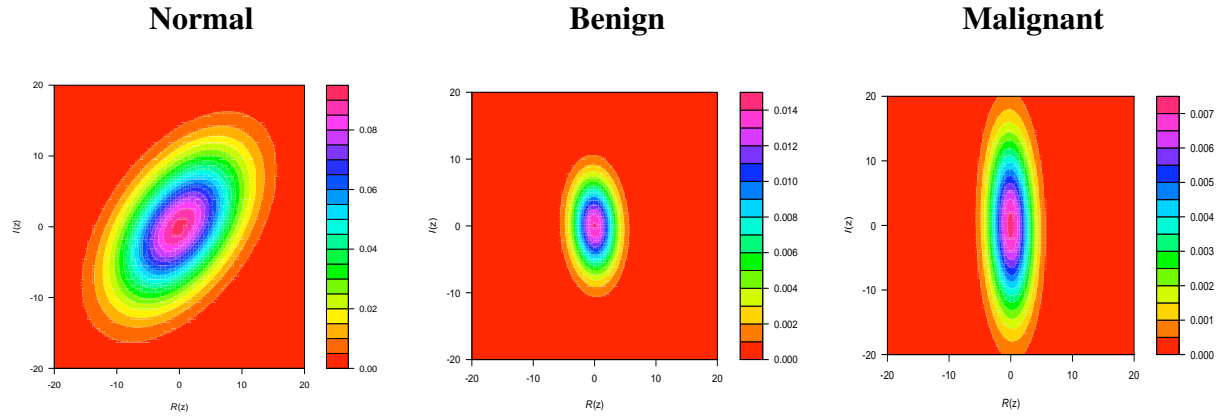


Figure 9.4: Bivariate Gaussian probability density functions estimated for adaptive image transform coefficients for the Pet hat filter, scale $a = 4$, location $(x, y) = (264, 102)$ for the normal, benign and malignant groups. This location is associated with amorphous scatter.

Parametric and non-parametric bivariate probability density functions were estimated for the coefficients of the adaptive image transform. Bivariate Gaussian probability density functions were fit to describe a large number of coefficient filter, scale and location indices. For instance, the Pet hat filter at scale $a = 4$ and location $(x, y) = (264, 102)$ corresponds to the amorphous scattering feature in the SAXS images and whose coefficients are fit using a bivariate Gaussian probability density function for each group. Figures 9.4 displays the bivariate Gaussian probability density function estimates for these adaptive image transform coefficients (Pet hat filter, scale $a = 4$, location $(x, y) = (264, 102)$) for the normal, benign and malignant groups. All probability density function estimates are all approximately centred on $(x, y) = (0, 0)$ but differ in their scale (variance) and orientation (covariance matrices). Inspection of the probability density function estimates suggests that in some cases adaptive image transform coefficients that are associated with amorphous scatter are in some instances reasonable at separating tissue groups and in some cases they are not. Relying on adaptive image transform coefficients associated with amorphous scatter alone may not produce a reliable classification of tissue group.

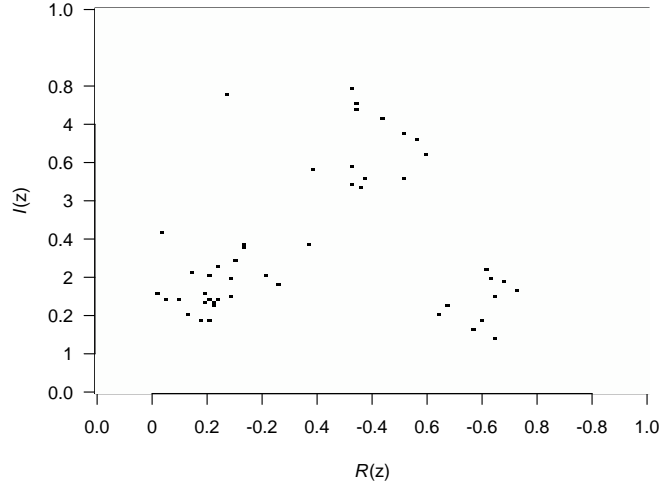
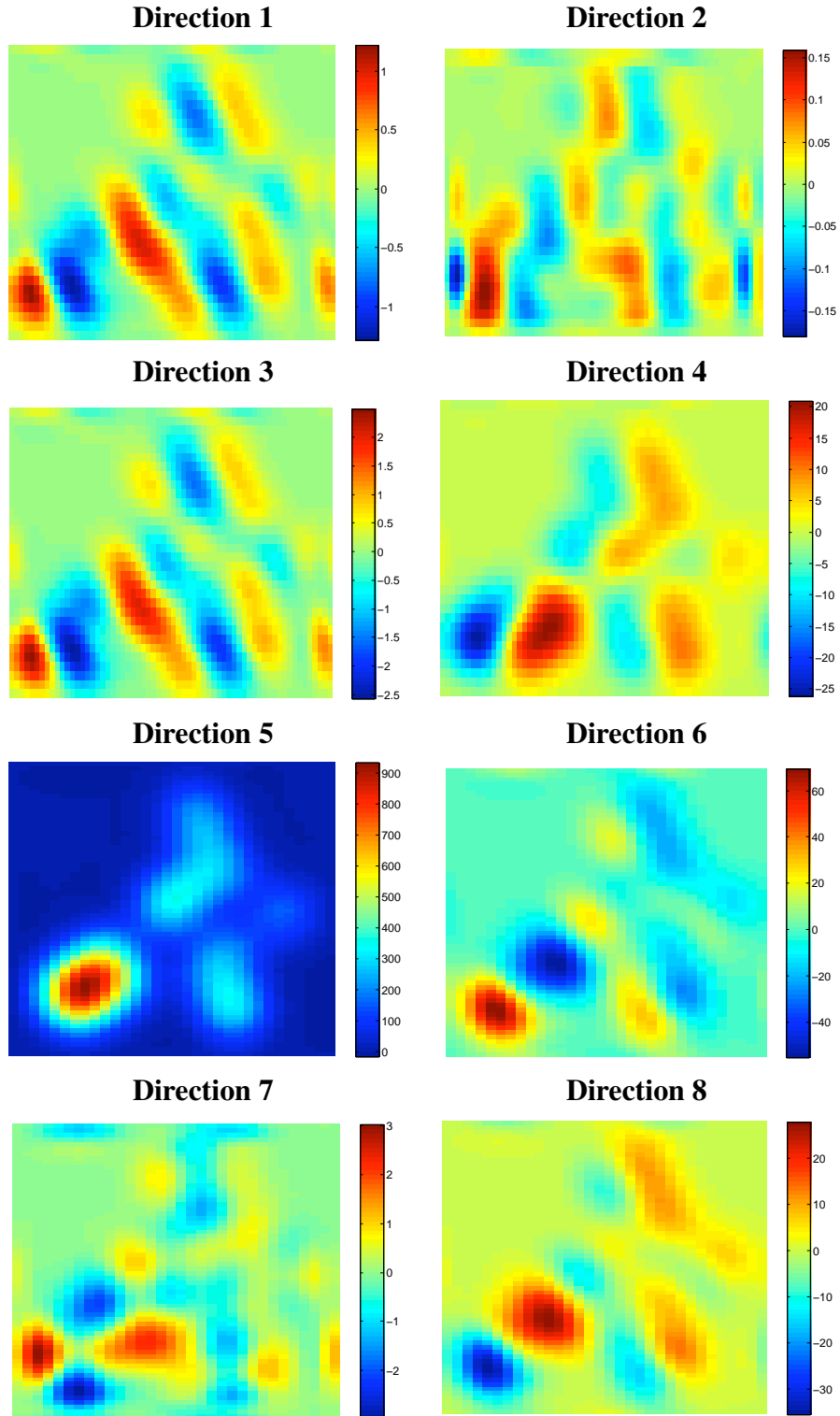


Figure 9.5: Raw histogram of adaptive image transform coefficients $z_{l,a,x,y}$ in the training data set for the malignant group, the Witch's Hat filter, scale $a = 2$, location $(x, y) = (256, 145)$.

The Mexican hat contourlet transform was also used to estimate select probability density functions non-parametrically. Recall that Section 8.2.4 introduced the Mexican hat contourlet transform, whilst Section 8.2.5 through to Section 8.2.8 developed methodology to estimate probability density functions by the modification of coefficient magnitudes. Non-parametric probability density function estimation using the Mexican hat contourlet transform was performed on approximately nine percent of adaptive image transform coefficients. The first step in this process was to estimate the raw histograms of adaptive image transform coefficient magnitude. An example raw histogram of adaptive image transform coefficients (for the malignant tissue group, the Witch's Hat filter, scale $a = 2$, location $(x, y) = (256, 145)$) is displayed in Figure 9.5. Take note of the three clusters of observations. This scale-location is of particular interest because it is associated with the location of the third-order meridional scattering ring in the original SAXS image, a feature studied extensively in previous works such as Round (2006) and Sidhu *et al* (2008).

Figure 9.6 displays select coefficient matrices of the corresponding Mexican hat contourlet transform (scale $j = 2$, for $l_2 = 8$ directions over all locations) of the raw histogram. Observe the vast differences in coefficient magnitudes between different directions. For example, compare the range of coefficient magnitudes in direction matrix $l_2 = 1$ (range of approximately $[-1, 1]$) with those in direction matrix $l_2 = 5$ (range of approximately $[0, 100]$). It would appear that direction matrix $l_2 = 5$ detects important structures in the raw histogram whilst the direction matrix $l_2 = 1$ appears to detect less structure.

Figure 9.6: Mexican Hat contourlet transform coefficients (scale $j = 2$, $l_2 = 8$ directions) of the raw histogram displayed in Figure 9.5.



The Mexican hat contourlet transform coefficients of the raw histogram were modified using the generalised extreme value distribution functions, $[\hat{P}(z, \hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1), \hat{P}(z, \hat{\mu}_2, \hat{\sigma}_2, \hat{\xi}_2), \dots, \hat{P}(z, \hat{\mu}_8, \hat{\sigma}_8, \hat{\xi}_8)]$ that were estimated using the methodology of Section 8.2.8.4. Figures 9.7 display the estimates of the generalised extreme value distribution functions for scales $j = (2, 4, 6)$.

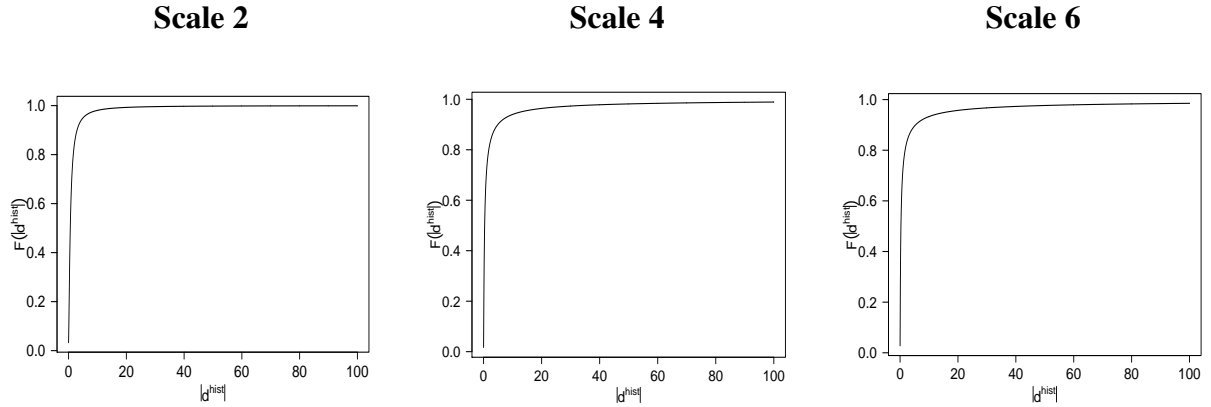


Figure 9.7: Generalised extreme value distribution weighting functions $[\hat{P}(|d^{\text{hist}}|, 0.4, 0.5, 0.7), \hat{P}(|d^{\text{hist}}|, 0.2, 0.3, 1.3), \hat{P}(|d^{\text{hist}}|, 0.2, 0.3, 1.5)]$. These functions are used to modify Mexican hat contourlet transform coefficient magnitude (at scales $j = 2, 4, 6$) prior to back transformation to yield the probability density function estimate.

All of the generalised extreme value distribution smoothing functions in Figure 9.7 modify the Mexican hat contourlet transform coefficient magnitudes by setting those below a certain value to zero and modifying those above this value. The threshold that governs the magnitude of the coefficients to set to zero is given by the μ_j parameter in each of the smoothing functions $\hat{P}(z, \mu_j, \sigma_j, \xi_j)$. Therefore for scale $j = 2$ all Mexican hat contourlet transform coefficients below $|d^{\text{hist}}| < 0.40$ in magnitude are set to zero. Similarly, those below 0.2 for scale $j = 4$ and 0.2 for scale $j = 6$ are set to zero. The Mexican hat contourlet transform coefficients above these magnitudes are modified according to a smooth trend. These smooth trends resemble step functions that has had its corner smoothed. The greatest amount of coefficient modulation occurs in the range $0.5 < |d^{\text{hist}}| < 20$ with approximately 10 to 20 percent reduction in magnitude being typical.

Figure 9.8: Probability Density Function Estimates (prior to normalisation) for the adaptive image transform of SAXS images for the Witch’s Hat filter, scale $a = 2$, location $(x, y) = (256, 145)$ in the perspective view. The red dots indicate the location of random samples from the *validation* data set.

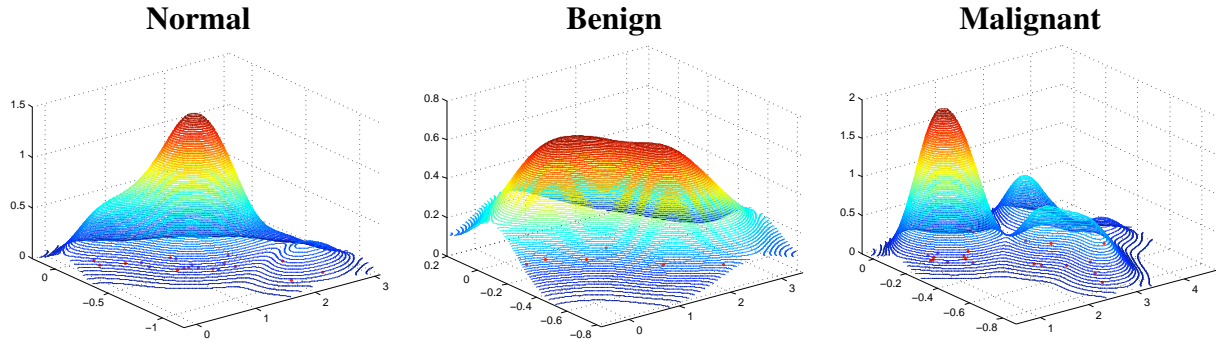
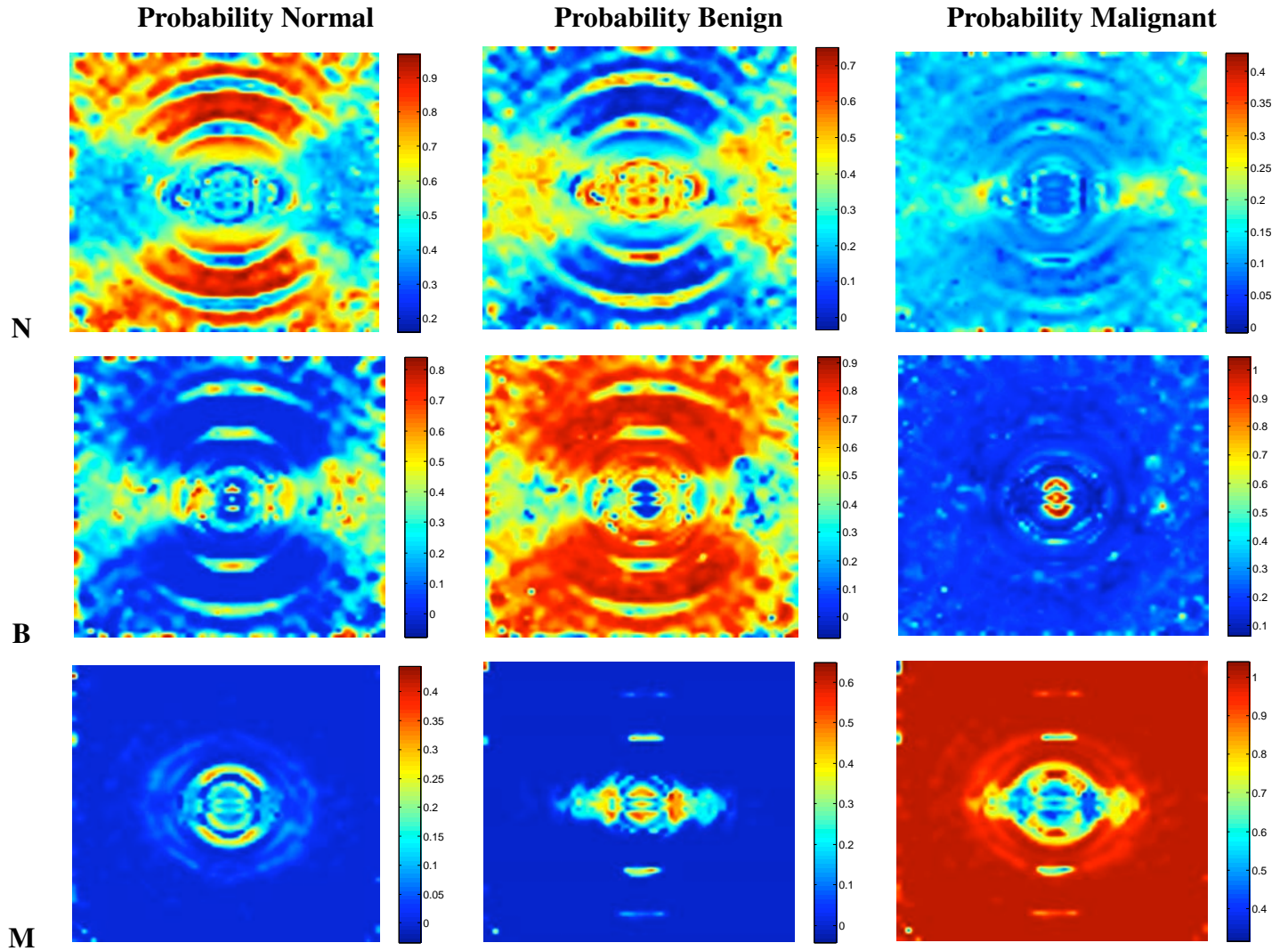


Figure 9.8 displays the non-parametric probability density function estimates of the adaptive image transform coefficients using Witch’s Hat filter at scale $a = 2$, location $(x, y) = (256, 145)$ for the normal, benign and malignant groups. These probability density function estimates were formed by constructing raw histograms of the adaptive image transform coefficients (using the training data set), taking the Mexican hat contourlet transform of each raw histogram, modifying the coefficient magnitude using the generalised extreme value function smoothing functions and then back-transforming. In fact the probability density function estimate for the malignant group used the raw histogram displayed in Figure 9.5.

Inspection of Figure 9.8 indicates some difference between these probability density functions between groups. A key difference between the probability density function estimates of the normal and malignant groups appears to be the greater number of modes (the peaks in the density function) which might indicate sub-populations within the data from the malignant tissue group. Note the overall difference in shape between the probability density function estimates between the three different tissue groups. The probability density function of the benign tissue group appears to be far more diffuse and non-centralised than either of the probability density functions associated with the normal or malignant tissue groups. The probability density function of the normal tissue group is centered on $[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})] = (1.4, -0.2)$ as compared to that of the malignant group which is centered at $[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})] = (0.9, -0.6)$. The red dots indicate the position of the coefficient magnitudes $[\Re(z_{l,a,x,y}), \Im(z_{l,a,x,y})]$ for a random sample of data from the *validation* data set. The coefficient magnitudes all fall within the non-zero regions of the probability density function estimate that corresponds to their group. This provides us with confidence that the probability density function estimates are a reasonable representation of adaptive image transform coefficient magnitude.

Despite the adaptive image transform coefficients associated with the third-order meridional scattering ring having a relatively high probability of misclassification, these coefficients still might convey useful diagnostic information and allow separation of tissue group to some extent. It must be kept in mind that Figure 9.8 displays probability density function estimates for one of the many possible filters, scales and locations of the adaptive image transform. Nonetheless this figure demonstrates the general trend observed when inspecting all of the probability density function estimates, which is that a certain degree of separation of the normal, benign and malignant tissue groups is possible using each adaptive image transform coefficient independently.

Figure 9.9: Probability estimates across locations for SAXS images (viewed at scale $a = 2$) of normal (N), benign (B) and malignant (M) breast tissue.



The adaptive image transform was taken of SAXS images from the validation data set, the real and imaginary components of each coefficient at each filter, scale and location were plugged into the bivariate probability density function estimates (from the previous step using the training data) for each group and Bayes' rule applied. For the j th SAXS image in the validation data set three arrays of probability were produced corresponding to the $\mathbf{P}_{j,\text{norm}}$, $\mathbf{P}_{j,\text{ben}}$ and $\mathbf{P}_{j,\text{mal}}$ arrays in Section 8.2.9. Estimates of the posterior probability for each group are displayed for three SAXS images in Figure 9.9, the scale is fixed at $a = 2$ and all locations are displayed. The probability estimates for only one scale are displayed because the $\mathbf{P}_{j,\text{norm}}$, $\mathbf{P}_{j,\text{ben}}$ and $\mathbf{P}_{j,\text{mal}}$ are quite large and space does not permit displaying the full arrays for each SAXS image in the validation data set.

The rows in Figure 9.9 correspond to the SAXS images, so that the symbol 'N' in the first row indicates the sample known to have a normal pathology, the symbol 'B' in the second row the benign pathology and the symbol 'M' in the third row the malignant pathology. The columns correspond to the probability estimates, the first column corresponds to probability of being normal tissue, the second column the probability of being benign tissue and the third column to the probability of being malignant tissue. Therefore scanning across the top row of Figure 9.9 reveals the location dependent probabilities for the SAXS image of the normal tissue sample.

It seems very likely that the sample would overall be classified as being normal healthy due to the large number of coefficients associated with a high probability of being normal. The scale-location dependent probability estimates have also provided insight into the most promising regions in the SAXS image for diagnosis. Observe that the greatest probabilities (of belonging to normal tissue) are in the regions between the third and fifth meridional scattering rings. This result supports the suspicions that it is the differences in the regions that are not readily visible to the eye in that are crucial to accurate diagnosis. The adaptive image transform appears to be a mathematical filter that is capable of detecting such differences.

The second row of Figure 9.9 is associated with a SAXS image from a benign sample. The amorphous scatter region again appears important but the specific regions are different from the normal sample. Also note that a large proportion of the coefficients also return a high probability of belonging to the benign group. There are some important exceptions, in particular compare the probability benign estimate with that of probability malignant. The central disk region in the SAXS image appears to be of importance, coefficients from this region have a very high probability of being incorrectly assigned to the malignant group. This highlights the problems associated with extracting a subset of coefficients for diagnosis. Incorrect selection will result in unnecessary errors in classification.

The bottom row of Figure 9.9 reveals that a great proportion of the adaptive image transform coefficients have a high probability of being assigned correctly to the malignant group. The amorphous scatter regions of the SAXS image have highest probability whilst the central disk, equatorial and meridional scattering rings regions has the lowest. In particular, the first-order equatorial scattering rings have (in general) a greater probability of belonging to the benign than the malignant groups. This result again highlights the importance of including all of the information in the SAXS image, selecting only a few coefficients may produce a model that correctly classifies one group at the expense of another.

9.3 Estimating the Overall Posterior Probability of a SAXS image belonging to each Group

The objective of this section was to estimate the overall posterior probabilities that SAXS images in an independent *test* data set belong to either the normal, benign or malignant groups. To achieve this objective the methodology of Section 8.3 was implemented. The arrays $\mathbf{P}_{j,\text{norm}}$, $\mathbf{P}_{j,\text{ben}}$ and $\mathbf{P}_{j,\text{mal}}$ from the validation data set were averaged according to equation (8.16) to produce the naive triplet of probabilities $[\hat{\text{Prob}}_{j,\text{norm}}, \hat{\text{Prob}}_{j,\text{ben}}, \hat{\text{Prob}}_{j,\text{mal}}]$. The naive probabilities were used to construct the raw histograms $[h_n(\hat{\text{Prob}}_{\text{norm}}), h_b(\hat{\text{Prob}}_{\text{ben}}), h_m(\hat{\text{Prob}}_{\text{mal}})]$, the Walsh wavelet packet transform was performed on these raw histograms, the coefficient magnitude modified and the inverse Walsh wavelet packet transform performed to yield the initial probability density function estimates $[\hat{p}(\hat{\text{Prob}}_{\text{norm}}), \hat{p}(\hat{\text{Prob}}_{\text{ben}}), \hat{p}(\hat{\text{Prob}}_{\text{mal}})]$. A separate *adjustment* SAXS image data set was then used refine these probability density function estimates to ensure the best classification performance. This refinement process was done according to the methodology specified in Section 8.3.5. The refined probability density function estimates were then used to classify the SAXS images in the test data set via Bayes' rule.

9.3.1 Histograms of Naive Probabilities Conditional on True Group

Figures 9.10 displays the histogram estimates of the naive probabilities as well as their corresponding smoothed probability density function estimates. Each of these histograms and probability density function estimates was estimated using samples that were of known pathology. For instance, the 'Prob N' histogram was constructed using naive probabilities from samples in the *validation* set known to be in the normal group. The 'Prob B' histogram was constructed from samples known to be in the benign group and the 'Prob M' histogram from samples known to be in the malignant group.

These naive estimators of the probability that a sample belongs to each group were inappropriate because they were produced by averaging of the probability arrays $\mathbf{P}_{j,\text{norm}}$, $\mathbf{P}_{j,\text{ben}}$ and $\mathbf{P}_{j,\text{mal}}$ for the $j = 1, \dots, N_{\text{val}}$ samples in the validation data set. This averaging assumes that the probability estimates at each filter, scale and location of the adaptive image transform coefficients are mutually exclusive. This need not be the case and indeed it seems unlikely as neighbouring adaptive image transform coefficients would be expected to have a certain amount of correlation. The Walsh wavelet packet transform was used to produce probability density function estimates (which are displayed on the bottom row of Figure 9.10) by retaining only select crystals and then back-transforming. A separate *adjustment* data set was used to calibrate these probability density function estimates in a way that maximised the correct number of classifications in each group. Considering all the naive probability estimates within a group and optimising the number of correct predictions on a separate data set allows for the refinement of the probability estimate that a sample belongs to a particular group.

Inspection of the probability density function estimates (along the bottom row of Figure 9.10) reveals that the Walsh wavelet packet method produces estimates that reasonably (visually) match well with the histograms. There are several spurious lobes such as the one on the interval $[0.0, 0.2]$ for the ‘Prob N’ probability density function estimate. One possibility is that the initial naive probability estimates are too crude a starting point in some cases, but even in that scenario the correction using an adjustment data set ensures that such artifacts are removed and therefore do not mislead classification. Another possibility is that the lobes are the result of working with a discrete data set. Walsh wavelet packet analysis functions belonging to a large scale might have overlapped with only a part of the non-zero count region of the raw histogram of the naive probability. The Walsh wavelet packet analysis functions would have then over-estimated the size of the ‘jump’ in the raw histogram counts. The coefficients produced by these analysis functions might have had a larger magnitude than those that would have been produced if there was a greater amount of data in the sparse regions of the histogram. Subsequent removal of these coefficients (by setting them to zero) using an *adjustment* data set might have optimised classification but has also induced oscillations in the resulting probability density function estimate.

This is because the spurious magnitude coefficients are more ‘significant’ to the analysis than need be and their removal omits low scale components. These low-frequency components are needed to cancel out the spurious oscillations of coefficient magnitude in the higher scales and so their inadvertent removal can result in probability density function estimates with spurious oscillations. This problem in the Walsh wavelet packet probability density function procedure can be reduced in many cases by including more data. Despite the presence of ‘lobes’ these probability density function estimates obtained were found to be sufficient for the application of breast cancer diagnosis.

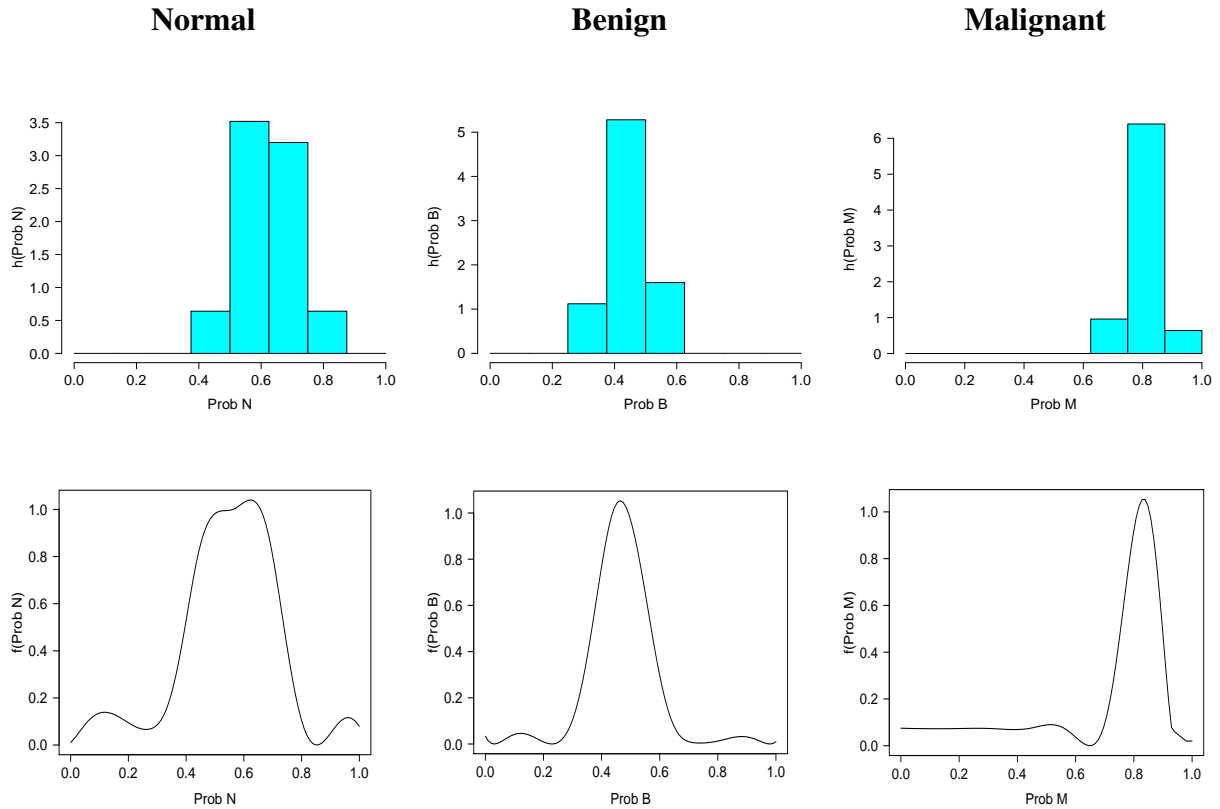


Figure 9.10: Estimation of the univariate probability density functions $[p(\hat{\text{Prob}}_n), p(\hat{\text{Prob}}_b), p(\hat{\text{Prob}}_m)]$. First row: histograms of the naive probabilities $[\hat{\text{Prob}}_{\text{norm}}, \hat{\text{Prob}}_{\text{ben}}, \hat{\text{Prob}}_{\text{mal}}]$ (denoted Prob N, Prob B and Prob M), second row: univariate probability density function estimates obtained using the Walsh wavelet packet basis and an *adjustment* data set.

9.4 Application of the Model to an Independent Test SAXS Image Data Set

The methodology described in Chapters 7 and 8 was applied to an independent *test data set*. The data is considered independent as the SAXS images involved were produced using tissue from a separate set of subjects and hence none of these images were used in prior model development. The adaptive image transform coefficients of the SAXS images in the test data set were found, the individual filter, scale and location probabilities estimated, the naive probabilities calculated and then adjusted using the probability density function estimates $[p(\hat{\text{Prob}}_n), p(\hat{\text{Prob}}_b), p(\hat{\text{Prob}}_m)]$. The probability that a sample (from the test data set) belongs to each group was calculated using these probability density function estimates at fixed values of ‘Prob N, Prob B, Prob M’. The sample is assigned to the group with the greatest magnitude of adjusted probability.

Note that the three adjusted probabilities that the sample belongs to each group need not sum to one. This is perfectly acceptable and the model was designed to produce estimated in this manner. Tissue samples from test data (those to be analysed in practice) are likely to contain a mixture of tissue from each group. A tissue sample need not belong to an exclusive tissue pathology. Each tissue sample is approximately ten milli-metres long and one milli-metre wide, it is reasonable to expect that different parts of the sample may contain different tissue pathologies. For instance, a situation that could arise when taking a biopsy that tissue is obtained that contains normal tissue in the lower five milli-metres and malignant tissue in the upper five milli-metres of the sample. Imaging with a SAXS camera would average out the signal from these two tissue components, the adaptive image transform and the related methodology is designed to extract both components of the signal from the SAXS image but when reporting the *overall* probability it would be incorrect to use methodology that assumes that the sample contains exclusively one pathology or the other. The probability estimation method by design takes this issue into account and therefore the adjusted probabilities need not sum to one.

Table 9.4 reports on the result of applying the SAXS image classification model that was developed in this thesis to an *independent* test data set. Classification was performed by assigning the sample to the group with the largest magnitude adjusted probability. The full set of probability estimates for the samples in this table are reported appendix A. The classification results were encouraging with only four out of thirty samples (of malignant pathology) being incorrectly classified as normal. The total accuracy was $86.7 \pm 3.3\%$, with the accuracy conditional that the samples were truly malignant being $60.0 \pm 10\%$. The accuracy of the model in correctly classifying malignant tissue samples was of concern and was investigated. The four samples that were incorrectly classified as being of normal pathology had a high proportion of normal healthy tissue present. This resulted in high levels of overall probability for both normal as well as malignant tissue. The high proportion of normal tissue in samples labeled as malignant may explain the reason for these classification errors and if all of the probability estimates are inspected rather than using hard classification then these borderline cases can be identified.

Table 9.4: Assessment of the SAXS Image Classification Model using an Independent Test Data Set.

		Classification		
		N	B	M
True	N	10	0	0
	B	0	10	0
	M	4	0	6

It must be stressed that ‘hard’ classification, assignment of a sample to a particular pathology is erroneous in this application. Table 9.4 is a guide to compare classification results of the model developed in this thesis to those of other researchers. Making informed decisions by inspecting the probability arrays such as those produced in Section 9.2 and examining all three of the overall probabilities such as those produced in this section is a sensible approach to the problem. If a sample has a high overall probability of being normal as well as a high overall probability (such as in the case of the samples that were incorrectly classified in the test data set) than a practitioner can inspect these probabilities and make a judgement out the need for the sample to undergo further scrutiny using other techniques. This approach appears to be a fair and sensible application of the SAXS image diagnostic model.

As a final assessment of the model that was developed in this thesis we compared our results to those of quadratic discriminant analysis models. The coefficients at location $(x, y) = (245, 145)$ of the Mexican hat wavelet transform were extracted for the thirty SAXS images in the test data set. Recall that a coefficient at this location is associated with third-order meridional scattering ring of the SAXS image. Therefore, the quadratic discriminant analysis models that were fitted could be considered to be a classical analysis that used coefficients detected by visual inspection from a fixed, off-the-shelf transform. Such models have been the focus of previous SAXS image analysis models such as those of Falzon *et al* (2006).

Tables 9.5 and 9.6 display the confusion matrices (assessed by leave-one out cross-validation) for the quadratic discriminant analysis models fit using Mexican hat wavelet transform coefficients. Table 9.5 displays the results for the coefficients extracted from scale $a = 2$, location $(x, y) = (245, 145)$, whilst Table 9.6 displays the results for scale $a = 4$, location $(x, y) = (245, 145)$.

Table 9.5: Classification of SAXS data using the Mexican Hat Wavelet Transform Coefficients at Scale $a = 2$, Location $(x, y) = (245, 145)$.

		Classification		
		N	B	M
True	N	8	0	2
	B	2	6	2
	M	5	4	1

Table 9.6: Classification of SAXS data using the Mexican Hat Wavelet Transform Coefficients at Scale $a = 4$, Location $(x, y) = (245, 145)$.

		Classification		
		N	B	M
True	N	8	0	2
	B	3	4	3
	M	4	4	2

Inspection of Tables 9.5 and 9.6 reveals that both quadratic discriminant analysis models perform poorly. The first classification model (at scale $a = 2$) has a total accuracy of $50.0 \pm 3.3\%$, note in particular that only one sample of malignant pathology is correctly classified. The second classification model (at scale $a = 4$) has a total accuracy of $46.6 \pm 3.3\%$, with only two samples of malignant pathology correctly classified. These models perform poorly compared to the adaptive image transform model developed in this thesis whose results are displayed in Table 9.4. The comparison of the results of these models reinforces the success of the methodology developed in this thesis.

9.5 Summary & Conclusions to Application of the Model

This chapter has involved the application of the models and methods developed in Chapters 7 and 8 to four SAXS image data sets. Good separation of the data was achieved in a manner that allowed insight into the actual differences between SAXS images produced by samples of different tissue pathology. Final model assessment using an independent test data set produced good results, but also highlighted the need to use all aspects of the model to make an informed decision regarding the state of the tissue.

Chapter 10

Summary, Conclusions and Future Directions

10.1 Summary

The central topic of this thesis has been the development of an improved diagnostic model of breast cancer using small-angle x-ray scattering images. In order to meet this objective a range of new image analysis and statistical modeling techniques were developed. Specific image analysis tools created include the adaptive image transform (Section 7.2), the Gamma, Mellin, Chebyshev, Witch's Hat and Zeta filter functions (Section 7.3) and a filter function selection algorithm based upon the probability of misclassification (Section 7.4). The statistical tools devised to model the adaptive transform coefficients of SAXS images included the XOR density estimation model (Section 8.2.2), the Mexican hat contourlet transform (Section 8.2.4), non-parametric probability density function estimation using extreme value theory (Section 8.2.8) and developing methods to smooth data defined on the interval $[0, 1]$ (Sections 8.3.3 and 8.3.5). These methods were applied to four SAXS image data sets in Chapter 9 to produce an accurate classifier of breast cancer.

10.2 Conclusion

Recall that the objectives of this research project were:

- a) To unite the capabilities of the adaptive transform in describing digital images and statistical inference within the framework of a mathematical model.
- b) To develop a unified, comprehensive, objective and semi-automated method with which to analyse SAXS images.
- c) To apply both the transform and the associated model to the diagnosis of breast cancer using SAXS images.

This thesis achieved its objectives:

- i) A model was developed that used the coefficients of an adaptive transform of SAXS images in a statistical model to classify tissue group.
- ii) The model is unified, it starts with the original SAXS image and in a series of logical steps produces three overall probabilities that the sample is one of three pathologies.
- iii) The model is comprehensive in that it considers a large amount of SAXS image information using a range of filter functions in order to search for any evidence (however) slight that the sample belongs to a particular tissue group.
- iv) The model is objective in the sense that the classification of group is based upon the probability estimates for which the methodology is explicitly defined. There is subjectivity in the interpretation of the results, in particular whether a sample needs further assessment but this is post implementation of the model.
- v) The model is almost fully automated, aside from pre-processing the SAXS images just need to be supplied to produce the probability estimates.
- vi) The adaptive image transform and the associated statistical model was applied to SAXS images of breast tissue.

A number of desirable outcomes were the result of achieving these objectives:

- a) A range of filter functions were developed to analyse SAXS images.
- b) An adaptive image transform was developed that allows a variety of filter functions to be used to analyse the same SAXS image, unlike the discrete wavelet transform which uses a fixed wavelet analysis function.
- c) The XOR model was developed as a method with which to store a large number of probability density functions in a manner that still retains the statistical properties of salient image features.
- d) The Mexican hat contourlet transform was developed in conjunction with coefficient thresholding using the generalised extreme value distribution functions to estimate bivariate probability density functions.
- e) Coefficient thresholding methodology using the Mexican hat contourlet transform/generalised extreme value distribution functions can be developed for any type of stochastic process that can be simulated. It need not be confined to the heteroskedastic noise $\mathcal{N}(0, \sigma_r^2)$, $\sigma_r^2 \sim U(0, 1)$ used in this thesis.
- f) The Walsh wavelet packet transform can be used to estimate probability density functions defined on the interval $[0, 1]$ in such a way that optimises classification performance.
- g) A general purpose transform and statistical methodology was developed that can be used in a range of other SAXS image classification scenarios. The methodology of Chapters 7 and 8 just need to be reapplied to the new data set.

A number of limitations exist in the model that has been developed:

- i) Estimation of the probabilities at each filter, scale and location was heavily reliant on the accurate estimation of the relevant probability density functions for each tissue group. Poor estimates of the probability density functions could result in poor estimates of the filter, scale and location probability.
- ii) A large number of Shapiro-Wilk tests need to be performed in the XOR model step, despite corrections ‘false discoveries’ can still result leading to the estimation of the probability density function non-parametrically. This increases the computational burden of the model.
- iii) Estimation of the univariate probability density functions using the Walsh wavelet packet transform produced ‘lobe’ artifacts.

Despite these limitations the overall results of the classification model when applied to an independent test data set were encouraging with twenty six out of thirty samples correctly identified. The ability to visualise the probability of each tissue group across filters, scales and locations was of great utility. It provided insight and understanding into those regions of the SAXS image most useful for identifying a particular tissue pathology. Furthermore it produced evidence that selecting a small subset of transform coefficients for diagnosis can give misleading results and incorrect classification. The XOR probability density function estimation model combined allowed for more compact representations of the probability density function estimates whose storage could have otherwise been prohibitive. Calculation of the probability of error at each filter, scale and location of the adaptive image transform revealed that the choice of filter function was critical to extracting the information in the SAXS image indicative of tissue type. It also revealed that the regions of the coefficient matrices associated with the third and fifth meridional scattering rings had a higher probability of error than other regions. This was a very important discovery as previous research such as that of Falzon *et al* (2006) focused on discrete wavelet transform coefficients associated with these scattering features.

10.3 Future Directions

This thesis is applied in nature with the main task being to produce an accurate classifier of breast tissue pathology using SAXS images. Image analysis and statistical methodology had to be developed in order to meet the challenges encountered and the results presented in this thesis present several avenues of further investigation. These avenues can be divided into several categories: image analysis, statistical and applied physics.

10.3.1 Image Analysis

Further applications of the adaptive image transform to other image data sets might prove useful. Applications such as the detection of micro-calcifications (which may indicate of breast cancer) in digital mammograms and the detection of objects in satellite imagery data are some of the possibilities. In each instance the filter functions can be custom-designed to detect the objects of interest.

Developing a general class of filter functions to apply to any image is also another possibility. This general class of functions would be designed to capture all of the image characteristics in such a way that it would be known that each filter captures a particular characteristic. For instance, filter function one in the library would be known to analyse the image for spatial frequency, filter function two would be known to analyse the image for transient events such as edges and so on. Whether such a general class of filter functions exists is (to the best of our knowledge) unknown and would need to be proved. Furthermore finding such a general class of filter functions could be challenging in practice.

The Mexican hat contourlet transform could also be applied to the analysis of images. It might be useful in the removal of image noise. The majority of image noise removal algorithms (based on the wavelet and contourlet transforms) assumes additive Gaussian noise. That is the stochastic component of the image model is sampled from a $\mathcal{N}(0, \sigma^2)$ distribution. This assumption is too restrictive in practice with a range of other models for the stochastic component being plausible. The joint Mexican hat contourlet transform/generalised extreme value distribution coefficient thresholding methodology can be modified to suit the problem at hand, which is an advantage over current image noise removal procedures.

10.3.2 Statistics

Non-parametric bivariate probability density function estimation based upon the Mexican hat contourlet transform deserves a thorough theoretical investigation to fully understand the statistical properties of the estimate. Comparisons of the efficiency of the estimator with those of other methods (such as kernels or mixtures of Gaussians) would be of interest and guide us to the suitability of the technique for general purpose bivariate probability density function estimation.

Creation of a non parametric density estimator that combines a range of basis functions might also be useful for complicated probability density functions. Such an estimator might be useful when the density function to be estimated contains both isolated peaks and broad, smooth trends at certain orientations (smooth curves). A two-dimensional wavelet transform would likely capture the peaks efficiently but not the smooth curves, whilst the Mexican hat contourlet transform would capture the smooth curves well but would average out the isolated peaks. A combined estimator would allow each transform to capture particular features in the probability density function and thereby offer a more competitive density estimation method.

Methods that allow functional data analysis of arrays of large dimension need to be developed. Such methods would have allowed us to estimate the overall probabilities that a sample belongs to each group direct from the probability arrays of filters, scales and locations. Models developed using the philosophy of functional data analysis might offer an opportunity to overcome the problem of having many predictors (individual filter, scale and location probabilities) with relatively few observations (SAXS images). The functional data analysis approach would negate the need to estimate the naive probabilities, thereby avoiding a loss of (potentially complex) spatial relationships amongst the filter, scale and location probabilities which might provide superior classification performance. Further exploration and development of statistical methods to perform functional data analysis on arrays of large dimension could offer further improvement in SAXS image diagnostic models.

Multiple statistical testing was used in the XOR model in order to determine between a non-parametric and parametric fit for each coefficient. The problems associated with multiple statistical testing were reduced by using corrections based upon the false discovery rate technique. Nonetheless, developing the theory and methods specifically for image analysis applications where correlations amongst false discoveries are likely is of interest and may yield superior results.

10.3.3 Applied Physics

Finally, there are many practical applications of the model in biophysical settings. The model could be applied to the analysis of SAXS images produced by other cancers or it could be refined to provide a more detailed breast cancer diagnosis such as predicting the exact pathology or stage of the disease. Physical scientists could use the results of the filter, scale and location probabilities to develop a deeper understanding of the differences in the structure of normal, benign and malignant breast tissue. Imaging cancer samples at synchrotron facilities might be a costly and impractical solution and there may well be a need to extend the model to allow analysis of SAXS images produced by cheaper and more compact x-ray sources.

The methodology of cancer diagnosis using SAXS imaging is continually being refined and further model development must be maintained to keep up with and guide future advances. The SAXS breast cancer diagnosis project has offered challenges to and motivated several advances in the field of statistical image analysis. At the same time the results of statistical image analysis have fed back into the physical sciences, allowing greater understanding of complex and demanding problems. We hope this thesis has served not only to develop a diagnostic algorithm but also to demonstrate a harmonious union between the physical and statistical sciences.

Appendices

Appendix A: Probability Estimates of the Test Data Set

	P_n	P_b	P_m
N	0.46	0.25	0.29
N	0.49	0.18	0.33
N	0.43	0.17	0.39
N	0.43	0.29	0.29
N	0.40	0.31	0.29
N	0.46	0.26	0.27
N	0.50	0.25	0.26
N	0.49	0.25	0.26
N	0.49	0.25	0.26
N	0.50	0.25	0.26
B	0.22	0.43	0.26
B	0.27	0.53	0.25
B	0.24	0.48	0.24
B	0.23	0.54	0.22
B	0.32	0.43	0.26
B	0.20	0.43	0.37
B	0.21	0.43	0.37
B	0.21	0.43	0.37
B	0.19	0.42	0.39
B	0.19	0.42	0.39
M	0.47	0.22	0.52
M	0.26	0.22	0.52
M	0.25	0.29	0.46
M	0.21	0.35	0.44
M	0.27	0.30	0.43
M	0.21	0.35	0.44
M	0.45	0.19	0.34 *
M	0.36	0.32	0.32 *
M	0.37	0.27	0.36 *
M	0.40	0.29	0.32 *

Posterior probability estimates of the model on independent test data (rounded to two decimal places). First column: true class, Second column: P_n probability estimate of sample being normal healthy tissue, Third column: P_b probability estimate of sample being benign tissue and fourth column: P_m probability estimate of being sample malignant pathology. The symbol ‘★’ indicates that the sample was incorrectly classified using ‘hard’ group assignment.

References

- [1] Abe S 2005 *Support Vector Machines for Pattern Classification*, Springer-Verlag London Limited, London, UK.
- [2] Abramowitz M & Stegun IA 1972 *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Dover Publications Inc., New York, New York, USA.
- [3] Acharya T & Ray AK 2005 *Image Processing*, John Wiley & Sons, Hoboken, New Jersey, USA.
- [4] Achim A, Bezerianos A & Tsakalides P 2001 ‘Novel Bayesian multiscale method for speckle reduction in medical ultrasound images’, *IEEE Transactions on Medical Imaging*, **20**(8): 772-783.
- [5] Addison PS 2002 *The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance*, Institute of Physics Publishing, Bristol, UK.
- [6] Agrawala AK (ed) 1977 *Machine Recognition of Patterns*, IEEE Press, New York, New York, USA.
- [7] Ahmed MN, Yamany SM, Mohamed N, Farag AA & Moriarty T 2002 ‘A modified fuzzy c-means algorithm for bias field estimation and segmentation of MRI data’, *IEEE Transactions on Medical Imaging*, **21**(3):193-199.
- [8] Alvarez L, Castaño C, Garc a M, Krissian K, Mazorra L, Salgado A, S  nchez J 2008 ‘Multi-channel satellite image analysis using a variational approach’, *Pure & Applied Geophysics*, **165**(6):1071-1093.

- [9] Alvarez L, Lions PL & Morel JM 1992 'Image selective smoothing and edge detection by nonlinear diffusion', *SIAM Journal of Numerical Analysis*, **29**(3):845-866.
- [10] Amenitsch H, Rappolt M, Laggner P, Bernstoff S, Moslinger R, Fleischmann E, Wagner T, Lax S, Petru E, Hudabiunigg K 1999 'Synchrotron x-ray study at Trieste: No correlation between breast cancer and hair structure', *Synchrotron Radiation News*, **12**(5): 32-34.
- [11] Amit Y 1994 'A nonlinear variational problem for image matching', *SIAM Journal on Scientific Computing*, **15**(1):207-224.
- [12] Antoine J-P, Carrette P, Murenzi R & Piette B 1993 'Image analysis with two-dimensional continuous wavelet transforms', *Signal Processing*, **31**(3): 241-272.
- [13] Antoine J-P & Murenzi R 1996 'Two-dimensional directional wavelets and the scale-angle representation', *Signal Processing*, **52**(3): 259-282.
- [14] Antoine J-P, Barachea D, Cesar RM & da Fontura Costa L 1997 'Shape characterisation with the wavelet transform', *Signal Processing*, **62**(3):265-290.
- [15] Antoine J-P, Murenzi R, Vandergheynst P & Twareque Ali S 2004 *Two-Dimensional Wavelets and their Relatives*, Cambridge University Press, Cambridge, United Kingdom.
- [16] Antoni M, Barlaud M, Mathieu P & Daubechies I 1992 'Image coding using the wavelet transform', *IEEE Transactions on Image Processing*, **1**(2): 205-220.
- [17] Arnéodo A, Decoster N & Roux SG 2000 'A wavelet-based method for multifractal image analysis: I. Methodology and test applications on isotropic and anisotropic random rough surfaces', *European Physical Journal B Condensed Matter Physics*, **15**(3): 567-600.
- [18] Atkinson AC, Donev AN & Tobias RD 2007, *Optimum Experimental Designs, with SAS*, Oxford University Press, New York, New York, USA.
- [19] Aubert G & Vese L 1997 'A variational method in image recovery', *SIAM Journal on Numerical Analysis*, **34**(5):1948-1979.
- [20] Ausloos M & Ivanova K 2002 'Multifractal nature of stock exchange prices', *Computer Physics Communications*, **147**(1-2): 582-585.

- [21] Ayed BA 2006 'Unsupervised variational image segmentation/classification using a Weibull observation model', *IEEE Transactions on Image Processing*, **15**(11):3431-3440.
- [22] Bailey SE & Lynch JM 2005 'Diagnostic differences in the mandibular P4 shape between neanderthals and anatomically modern humans', *American Journal of Physical Anthropology*, **126**:268-277.
- [23] Bamberger RH & Smith MJT 1992 'A filter bank for the directional decomposition of images: theory and design', *IEEE Transactions on Signal Processing*, **40**(4):882-893.
- [24] Bankmans IN (ed) 2000 *Handbook of Medical Imaging and Analysis*, (2nd edn), Academic Press, San Diego, California, USA.
- [25] Barnsley MF 1988 *Fractals everywhere*, (2nd edn), Academic Press, New York, New York, USA.
- [26] Barnsley MF, Hurd LP & Anson LF 1993 *Fractal Image Compression*, AK Peters, Wellesley, Massachusetts, USA.
- [27] Batchelar DL & Cunningham IA 2002 'Material-specific analysis using coherent scatter imaging', *Medical Physics*, **29**(8): 1651-1660.
- [28] Baydush AH, Bowsher JE, Laading JK & Floyd Jr CE 1997 'Improved Bayesian image estimation for digital chest radiography', *Medical Physics*, **24**:539-546.
- [29] Beirlant J, Goegebeur Y, Segers J & Teugels J 2004 *Statistics of Extremes: Theory and Applications*, Wiley, Hoboken, New Jersey, USA.
- [30] Benjamini Y & Hochberg Y 1995 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *Journal of the Royal Statistical Society Series B*, **57**(1):289-300.
- [31] Bergeaud F & Mallat SG 1995 'Matching pursuit of images', *Proceedings of the SPIE*, **2491**: 2-13.
- [32] Berger J 1985 *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, New York, USA.

- [33] Berlinet A & Devroye L 1994 'A comparison of kernel density estimates', Publications de l'Institut de Statistique de l'Universite de Paris, **38**(3):3-59. Available at <http://cg.scs.carleton.ca/luc/devs.html>
- [34] Bertalmio M, Sapiro G, Caselles V & Ballester C 2000 'Image inpainting', *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, ACM Press/Addison-Wesley Publishing Company, New York, New York, USA, pp. 417-424.
- [35] Bertelli L, Sumengen B, Manjunath BS, Gibou F 2008 'A variational framework for multiregion pairwise-similarity-based image segmentation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **30**(8):1400-1414.
- [36] Bertero M & Bocacci P 1998 *Introduction to Inverse Problems in Imaging*, Institute of Physics Publishing, Bristol, UK.
- [37] Besag J 1974 'Spatial interaction and the statistical analysis of lattice systems', *Journal of the Royal Statistical Society Series B*, **36**:192-226.
- [38] Betal D, Roberts N & Whitehouse GH 1997 'Segmentation and numerical analysis of microcalcifications on mammograms using mathematical morphology', *The British Journal of Radiology*, **70**: 903-917.
- [39] Beucher S 1994 'Watershed, hierarchical segmentation and the waterfall algorithm', *Proceedings Mathematical morphology and its application to image processing*, Kluwer Academic, Netherlands. pp. 69-76.
- [40] Bezdek JC & Pal SK 1992 *Fuzzy Models for Pattern Recognition: Methods that Search for Structure in Data*, IEEE Press, New York, New York, USA.
- [41] Bezdek JC, Hall LO, Clark MC, Goldgof DB & Clarke LP 1997 'Medical image analysis with fuzzy models', *Statistical Methods in Medical Research*, **6**(3):191-214.
- [42] Bezdek JC, Keller J, Krisnapuram R & Pal NR 2005 *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Springer, New York, New York, USA.
- [43] Bloch I, Martino G & Petrosino A 2006 'A fuzzy mathematical morphology approach to multiseeded image segmentation', *Lecture Notes in Computer Science*, **3849**:362-368.

- [44] Blomgren P, Chan TF, Mulet P & Wong CK 1997 'Total variation image restoration: numerical methods and extensions', *Proceedings, 1997 International Conference on Image Processing*, **3**:384-387.
- [45] Blomgren P, Chan TF, Mulet P, Verse L & Wan WL 'Variational PDE models and methods for image processing' in Griffiths DF & Watson GA 1999 *Numerical Analysis*, Chapman & Hall/CRC, Boca Raton, Florida, USA.
- [46] Biernacki C, Celeux G, Govaert G & Langrognet F 2005 'Model-based cluster and discriminant analysis with the MIXMOD software', *Computational Statistics and Data Analysis*, **51**(2):587-600.
- [47] Bigi A & Roveri N 1996 'Fibre Diffraction:Collagen' in *Handbook on Synchrotron Radiation*, Ebashi S, Koch M & Rubenstein E (eds), Elsevier, **4**:199-239.
- [48] Binaghi E, Della Ventura A, Rampini A & Schettini R 2007 'Fuzzy reasoning approach to similarity evaluation in image analysis', *International Journal of Intelligent Systems*, **8**(7):749-769.
- [49] Bishop CM 1995 *Neural Networks for Pattern Recognition*, Oxford University Press, New York, New York, USA.
- [50] Boas ML 2006 *Mathematical Methods in the Physical Sciences* (3rd edn), John Wiley and Sons, Hoboken, New Jersey, USA.
- [51] Boles WW & Boashash B 1998 'A human identification technique using images of the iris and wavelet transform', *IEEE Transactions on Signal Processing*, **46**(4): 1185-1188.
- [52] Bookstein FL 1991 *Morpometric Tools for Landmarks Data: Geometry and Biology*, Cambridge University Press, Cambridge, United Kingdom.
- [53] Bovik AC '2.3: Basic Tools for Image Fourier Analysis' in Bovik AC (ed) 2005 '*Handbook of Image and Video Processing*', (2nd edn), Elsevier Academic Press, Amsterdam, The Netherlands.
- [54] Bovik AC (ed) 2005 '*Handbook of Image and Video Processing*', (2nd edn), Elsevier Academic Press, Amsterdam, The Netherlands.

- [55] Branner B ‘The Mandelbrot set’ in Devaney RL & Keen L (eds) 1988, *Chaos and Fractals. The Mathematics Behind Computer Graphics*, American Mathematical Society, Providence, Rhode Island, USA.
- [56] Breiman L, Friedman JH, Olshen RA & Stone CJ 1984 *Classification and Regression Trees*, Wadsworth & Brooks/Cole, Monterey, California, USA.
- [57] Breiman L 2001 ‘Random Forests’, *Machine Learning*, **45**(1):5-32.
- [58] Bromiley PA, Thacker NA, Scott MLJ, Pokrić M, Lacey AJ & Cootes TF 2003 ‘Bayesian and non-Bayesian probabilistic models for medical image analysis’, *Image and Vision Computing*, **21**(10): 851-864.
- [59] Briki F, Busson B, Salicrú B, Esteve F & Doucet J 1999 ‘Breast cancer diagnosis using hair’, *Nature*, **400**: 236.
- [60] Brillouin L 1963 *Science and Information Theory*, (2nd edn), Academic Press, New York, New York, USA.
- [61] Brook A, Kimmel R & Sochen NA 2003 ‘Variational restoration and edge detection for color images’, *Journal of Mathematical Imaging and Vision*, **18**(3):247-268.
- [62] Brunk HD 1978 ‘Univariate density estimation by orthogonal series’, *Biometrika*, **65**(3):521-528.
- [63] Buntine W 1992 ‘Learning classification trees’, *Statistics and Computing*, **2**(2):63-73.
- [64] Burt P & Adelson E 1983 ‘The Laplacian Pyramid as a Compact Image Code’, *IEEE Transactions on Communications*, **31**(4):532-540.
- [65] Bushberg JT, Seibert JA, Leidholdt Jr., EM, Boone, JM 2002, *The Essential Physics of Medical Imaging* (2nd Edn), Lippincott Williams & Wilkins, Philadelphia, USA.
- [66] Butler S, Webb GI & Lewis RA 2003 ‘A case study in feature invention for breast cancer diagnosis using x-ray scatter images’, *Lecture Notes in Computer Science*, **2903**: 677-685.
- [67] Candès EJ & Donoho DL 2000 ‘Curvelets: A surprisingly effective nonadaptive representation for objects with edges’, Technical Report, Department of Statistics, Stanford University, California, USA.

- [68] Candès EJ & Donoho DL 2003 'New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities', *Communications on Pure and Applied Mathematics*, **57**(2): 219-266.
- [69] Candès EJ, Demanet L, Donoho D & Ying L 2006 'Fast discrete curvelet transforms', *Multiscale Modeling and Simulation*, **5**(3): 861-899.
- [70] Castro CRF, Barroso RC, Anjos MJ, Lopes RT & Braz D 2004 'Coherent scattering characteristics of normal and pathological breast human tissues', *Radiation Physics and Chemistry*, **71**(3-4): 649-651.
- [71] Castro CRF, Barroso RC, de Oliveira LF & Lopes RT 2005 'Coherent scattering x-ray imaging at the Brazilian national synchrotron laboratory: preliminary breast images', *Nuclear Instruments and Methods in Physics Research A*, **548**(1-2): 116-122.
- [72] Castro CRF, Barroso RC & Lopes RT 2005 'Scattering signatures for some human tissues using synchrotron radiation', *X-ray Spectrometry*, **34**(6): 477-480.
- [73] Catte F, Lions PL, Morel JN & Coll T 1992 'Image selective smoothing and edge detection by nonlinear diffusion', *Numerical Analysis*, **29**(1):182-193.
- [74] Chambers JM, Cleveland WS, Kleiner B & Tukey PA 1983 *Graphical Methods for Data Analysis*, Wadsworth, Belmont, California, USA.
- [75] Chambolle A 1995 'Image segmentation by variational methods: Mumford and Shah functional and the discrete approximations', *SIAM Journal of Applied Mathematics*, **55**(3):827-863.
- [76] Chan FHY, Lam FK & Zhu H 1998 'Adaptive thresholding by variational method', *IEEE Transactions on Image Processing*, **7**(3):468-473.
- [77] Chan T, Marquina A & Mulet P 2000 'High-order total variation-based image restoration', *SIAM Journal on Scientific Computing*, **22**(4):503-516.
- [78] Chan T & Shen J 2001a 'Nontexture inpainting by curvature driven diffusion', *Journal of Visual Communication and Image Representation*, **12**(4):436-449.
- [79] Chan T & Shen J 2001b 'Variational restoration of nonflat image features: models and algorithms', *SIAM Journal on Applied Mathematics*, **61**(4):1338-1361.

- [80] Chan T & Shen J 2002 'Mathematical models for local nontexture inpaintings', *SIAM Journal of Applied Mathematics*, **62**(3):1019-1043.
- [81] Chan T & Shen J 2005a 'Image processing and analysis: variational, PDE, wavelet & stochastic methods', *Society for Industrial and Applied Mathematics*, Philadelphia, Pennsylvania, USA.
- [82] Chan T & Shen J 2005b 'Variational image inpainting', *Communications on Pure and Applied Mathematics*, **58**(5):579-619.
- [83] Chang SG, Yu B & Vertelli M 2000 'Adaptive wavelet thresholding for image denoising and compression', *IEEE Transactions on Image Processing*, **9**(4): 1532-1546.
- [84] Changizi V, Kheradmand AA & Oghabian MA 2008 'Application of small-angle x-ray scattering for differentiation among breast tumours', *IFMBE Proceedings*, **33**(1): 19-23.
- [85] Changizi V, Oghabian MA, Speller R, Sarkar S & Kheradmand AA 2005 'Application of Small Angle X-ray Scattering (SAXS) for Differentiation between Normal and Cancerous Breast Tissue', *International Journal of Medical Science*, **2**(3): 118-121.
- [86] Chen C, Ouyang X, Wang WH & Hu X 1992 'Improvement of medical images using Bayesian processing', *Proceedings of the SPIE*, **1652**:489-491.
- [87] Chen Y, Vemuri BC & Wang L 2000 'Image denoising and segmentation via nonlinear diffusion', *Computers and Mathematics with Applications*, **39**(5-6):131-149.
- [88] Chen SS, Donoho DL & Saunders MA 2001 'Atomic Decomposition by Basis Pursuit', *SIAM Review*, **43**(1): 129-159.
- [89] Cheng B & Titterton D 1994 'Neural networks: a review from a statistical perspective', *Statistical Science*, **9**:2-54.
- [90] Chi Z, Yan H & Pham T 1996 *Fuzzy Algorithms with Applications to Image Processing and Pattern Recognition*, World Scientific, Singapore.
- [91] Chu B, Fong D & Hsaio BS 1999 'Hair test results at the advanced polymers beamline (X27C) at the NSLS', *Synchrotron Radiation News*, **12**(5): 36.

- [92] Cohen A, Daubechies I & Feauveau J 1992 'Bi-orthogonal bases of compactly supported wavelets', *Communications on Pure and Applied Mathematics*, **45**(5): 485-560.
- [93] Cohen A, Daubechies I & Vial P 1993 'Wavelets on the interval and fast wavelet transforms', *Applied Computational and Harmonic Analysis*, **1**(1):54-81.
- [94] Coifman RR & Wickerhauser MV 1992 'Entropy-based algorithms for best basis selection', *IEEE Transactions on Information Theory*, **38**(2):713-718.
- [95] Combes J-M, Grossmann A & Tchamitchian P (eds) 1989 *Wavelets: Time-Frequency Methods and Phase Space*, Springer-Verlag, Berlin, Germany.
- [96] Cootes TF, Taylor CJ, Cooper DH & Graham J 1995 'Active Shape Models-Their Training and Application', *Computer Vision and Image Understanding*, **61**(11):38-59.
- [97] Costaridou L (ed) 2005 *Medical Image Analysis Methods*, Taylor & Francis, Boca Raton, Florida, USA.
- [98] Cover T & Hart P 1967 'Nearest neighbor pattern classification', *IEEE Transactions on Information Theory*, **13**(1):21-27.
- [99] Cremers D 2002 *Statistical Shape Knowledge in Variational Image Segmentation*, Doctoral Thesis, Department of Mathematics and Computer Science, University of Mannheim, Mannheim, Germany.
- [100] Cremers D, Kohlberger T & Schnörr C 2003 'Shape statistics in kernel space for variational image segmentation', *Pattern Recognition*, **36**:1929-1943.
- [101] Cristianini N & Shawe-Taylor J 1999 *An Introduction to Support Vector Machines: and other kernel-based learning methods*, Cambridge University Press, New York, New York, USA.
- [102] Cross SS 1997 'Fractals in pathology', *Journal of Pathology*, **182**:1-8.
- [103] Cunningham CC & Anthony D 1993 'Image deconvolution of extended objects: A comparison of the inverse Fourier and the Lucy techniques', *Icarus*, **102**(2):307-315.
- [104] da Fontoura Costa L & Cesar Jr RM 2001 *Shape Analysis and Classification: Theory and Practice*, CRC Press, Boca Raton, Florida, USA.

- [105] Daresbury Synchrotron Radiation Source 2009, 'BSL Manual', Retrieved 30 March 2009 from <http://www.srs.dl.ac.uk/ncd/computing/manual.bsl.html>
- [106] Dasarathy BV (ed) 1991 *Nearest Neighbour (NN) Norms: NN Pattern Classification Techniques*, IEEE Computer Society Press, Los Alamitos, California, USA.
- [107] Daubechies I 1992 *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA.
- [108] Daubechies I 1993 'Orthonormal bases of compactly supported wavelets II. Variations on a theme', *SIAM Journal on Mathematical Analysis*, **24**(2):499-519.
- [109] Daubechies I & Teschke G 2005 'Variational image restoration by means of wavelets: simultaneous decomposition, restoration, deblurring and denoising', *Applied and Computational Harmonic Analysis*, **19**(1):1-16.
- [110] Davatzikos C, Tao X & Shen D 2003 'Hierarchical active shape models using the wavelet transform', *IEEE Transactions on Medical Imaging*, **22**(3):414-423.
- [111] Davison AC 2003 *Statistical Models*, Cambridge University Press, New York, New York, USA.
- [112] DeCastro E & Morandi C 1987 'Registration of translated and rotated images using finite Fourier transforms', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-**9**(5):700-703.
- [113] De Felici M, Felici R, Ferrero C, Tartani A, Gambaccini M & Finet S 2008 'Structural characterisation of human cerebral myelin sheath by small-angle x-ray scattering', *Physics in Medicine and Biology*, **53**: 5675-5688.
- [114] Devaney RL & Keen L (eds) 1988, *Chaos and Fractals. The Mathematics Behind Computer Graphics*, American Mathematical Society, Providence, Rhode Island, USA.
- [115] De Ville V, Nachtegaal M, Van der Weken D, Kerre EE, Phillips W & Lemahieu I 2003 'Noise reduction by fuzzy image filtering', *IEEE Transactions on Fuzzy Systems*, **11**(4):429-436.
- [116] DeVore RA, Jawerth B & Lucier BJ 1992 'Image compression through wavelet transform coding', *IEEE Transactions on Information Theory*, **38**(2): 719-746.

- [117] Devroye L & Györfi L 1985 *Nonparametric Density Estimation: The L-1 View*, Wiley, New York, New York, USA.
- [118] Do MN & Vertelli M 2002 'Contourlets: A new directional multiresolution image representation', *Asilomar conference on signals, systems and computers*, **1**: 497-504.
- [119] Do MN & Vertelli M 2003 'Contourlets' pages 84-107 in *Beyond Wavelets* (1st Ed), Welland GV (ed), Academic Press, San Diego, California, USA.
- [120] Do MN & Vertelli M 2006 'The contourlet transform: An efficient directional multiresolution image representation', *Official Journal - European Union Information and Notices C*, **49**(27A):2091-2106.
- [121] Donoho DL 1993 'Nonlinear wavelet methods for the recovery of signals, densities and spectra from indirect and noisy data', *Proceedings of Symposia in Applied Mathematics*, **47**:173-205, American Mathematical Society.
- [122] Donoho DL, Johnstone IM, Kerkycharian G & Picard D 1996 'Density estimation by wavelet thresholding', *The Annals of Statistics*, **24**(2):508-539.
- [123] Donoho DL 2001 'Sparse Components of Images and Optimal Atomic Decompositions', *Constructive Approximation*, **17**(3):353-382.
- [124] Dougherty G 2001 'A comparison of the texture of computed tomography and projection radiographic images of vertebral trabecular bone using fractal signature and lacunarity', *Medical Engineering and Physics*, **23**: 313-321.
- [125] Dougherty ER, Astola JT & Egiazarian KO (eds) 2003 *Image Processing: Algorithms and Systems II*, SPIE-The International Society for Optical Engineering, Santa Clara, California, USA.
- [126] Dougherty ER & Lotufo RA 2003 'Hands on Morphological Image Processing', *SPIE Optical Engineering Press*, Bellingham, Washington, USA.
- [127] Drokse M & Rumpf M 2004 'A variational approach to nonrigid morphological image registration', *SIAM Journal of Applied Mathematics*, **64**(2):668-687.
- [128] Dryden IL & Mardia KV 1998 *Statistical Shape Analysis*, John Wiley & Sons, Chichester, New York, New York, USA.

- [129] Dugad R, Ratakonda K & Ahuja N 1998 'A new wavelet-based scheme for watermarking images', *ICIP 98, Proceedings 1998 International Conference on Image Processing*, **2**:419-423.
- [130] Dupuis P, Grenander U & Miller MI 1998 'Variational problems on flows of diffeomorphisms for image matching', *Quarterly of Applied Mathematics*, **56**(3):587-600.
- [131] Dutilleul P 'An implementation of the "algorithme à trous" to compute the wavelet transform'. In Combes J-M, Grossmann A & Tchamitchian P (eds) *Wavelets: Time-Frequency Methods and Phase Space*, 1989, Inverse Problems and Theoretical Imaging, pages 298-304, Springer-Verlag, Berlin, Germany.
- [132] , Ebashi S, Koch M & Rubenstein E (eds) 1996 *Handbook on Synchrotron Radiation*, Elsevier, Academic Press, Amsterdam, The Netherlands.
- [133] Efron B & Tibshirani RJ 1993 *An Introduction to the Bootstrap*, Chapman & Hall, New York, New York, USA.
- [134] Ehrhardt J, Schmidt-Richberg A & Handels H 2008 'Simultaneous segmentation and motion estimation in 4D-CT-data using a variational approach', *Proceedings SPIE*, **6914**(3):6914-6937.
- [135] Einstein AJ, Wu H-S & Gil J 1998 'Self-Affinity and lacunarity of chromatin texture in benign and malignant breast epithelial cell nuclei', *Physical Review Letters*, **80**:397-400.
- [136] Elshemey WM, Elsayed AA & El-Lakkani A 1999 'Characteristics of low-angle x-ray scattering from some biological samples', *Physics in Medicine and Biology*, **44**: 2907-2915.
- [137] Elshemey WM, Desouky OS, Mohammed MS, Elsayed AA & El-houseini ME 2003 'Characterization of cirrhosis and hepatocellular carcinoma using low-angle x-ray scattering signatures of serum', *Physics in Medicine and Biology*, **48**: N239-N246.
- [138] Engelberg S 2008 'Edge detection using Fourier coefficients', *American Mathematical Monthly*, **115**(6):499-513.
- [139] Erdélyi A, Magnus W, Oberhettinger F & Tricomi FG 1954, *Tables of Integral Transforms*, Volume I, McGraw-Hill, New York, New York, USA.

- [140] Erickson CM 2005 *Automated Detection of Breast Cancer using SAXS data and Wavelet Features*, Master of Science Thesis, Biomedical Engineering Department, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
- [141] Esedoglu S & Shen J 2002 'Digital-inpainting based on the Mumford-Shah-Euler image model', *European Journal of Applied Mathematics*, **13**(4):353-370.
- [142] Eslami R & Radha H 2003 'The contourlet transform for image denoising using cycle spinning', *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, 2003.*, **2**:1982-1986.
- [143] Evans SH, Bradley DA, Dance DR, Bateman JE & Jones CH 1991 'Measurement of small-angle photon scattering for some breast tissues and tissue substitute materials', *Physics in Medicine and Biology*, **36**: 7-18.
- [144] Evans DGR, Howell A, Hasnain SS, Grossmann JG 2001 'Science or black magic?', *Journal Medical Genetics*, **38**: e16.
- [145] Falzon G, Pearson S, Murison R, Hall C, Siu K, Evans A, Rogers K & Lewis R 2006 'Wavelet-based feature extraction applied to small-angle x-ray scattering patterns from breast tissue: a tool for differentiating between tissue types', *Physics in Medicine and Biology*, **51**: 2465-2477.
- [146] Falzon G, Pearson S, Murison R, Hall C, Siu K, Round A, Scültke E, Kaye AH & Lewis R 2007 'Myelin structure is a key difference in the x-ray scattering signature between meningioma, schwannoma and glioblastoma multiforme', *Physics in Medicine and Biology*, **52**: 6543-6553.
- [147] Family F & Vicsek T (eds) 1991 *Dynamics of Fractal Surfaces*, World Scientific Publishing, Singapore.
- [148] Farag AA, Ahmed MN, El-Baz A & Hassan H 'Advanced segmentation techniques' in Suri JS, Wilson DL & Laxminarayan S 2005 *Handbook of Biomedical Image Analysis: Volume I Segmentation Models Part A*, Kluwer Academic/Plenum Publishers, New York, New York, USA.
- [149] Faugeras O, Adde G, Charpiat G, chef d'Hotel C, Clerc M, Deneux T, Deriche R, Hermosillo G, Keriven R, Kornprobst P, Kybic J, Lenglet C, Lopez-Perez L, Papadopoulos

- T, Pans J-P, Segonne F, Thirion B, Tschumperlé D, Viéville T & Wotawa N 2004 'Variational, geometric, and statistical methods for modeling brain anatomy and function', *NeuroImage*, **23**(1):S46-S55.
- [150] Feder J 1988 *Fractals*, Pergamon, New York, New York, USA.
- [151] Feldmann A, Gilbert AC & Willinger W 1998 'Data networks as cascades: investigating the multifractal nature of internet WAN traffic', *ACM SIGCOMM Computer Communication Review*, **28**(4):42-45.
- [152] Feigin LA & Svergun DI 1987 *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*, Plenum Press, New York, New York, USA.
- [153] Fernández M, Keyriläinen J, Serimaa R, Torkkeli M, Karjalainen-Lindsberg M-L, Tenhunen M, Thomlinson W, Urban V, Suortti P 2002 'Small-angle x-ray scattering of human breast tissue samples', *Physics in Medicine and Biology*, **47**: 577-592.
- [154] Fernández M, Keyriläinen J, Karjalainen-Lindsberg M-L, Fiedler S, Suortti P 2004 'Human breast tissue characterisation with small-angle x-ray scattering', *Spectroscopy*, **18**: 167-176
- [155] Fernández M, Keyriläinen J, Serimaa R, Torkkeli M, Karjalainen-Lindsberg M-L, Leidenius M, von Smitten K, Tenhunen M, Fiedler S, Bravin A, Weiss TM & Suortti P 2005 'Human breast cancer *in vitro*: matching histo-pathology with small-angle x-ray scattering and diffraction-enhanced imaging', *Physics in Medicine and Biology*, **50**: 2991-3006.
- [156] Ferrar WL 1927 'On the consistency of cardinal function interpolation', *Proceedings of the Royal Society of Edinburgh*, **47**: 230-242.
- [157] Fieguth P & Zhang J 'Random Field Models' in Bovik A (ed) 2005 *Handbook of Image and Video Processing*, Elsevier Academic Press, Amsterdam, The Netherlands.
- [158] Fischl B & Dale AM 2000 'Measuring the thickness of human cerebral cortex from magnetic resonance images', *Proceedings of the National Academy of Sciences*, **97**(20):11050-11055.
- [159] Fisher RA & Tippett LHC 1928 'Limiting forms of the frequency distribution of the largest or smallest member of a sample', *Mathematical Proceedings of the Cambridge Philosophical Society*, **24**(2):180-190.

- [160] Fisher RA 1936 'The use of multiple measurements in taxonomic problems', *Annals of Eugenics*, **7**:179-188.
- [161] Fisher Y (ed) 1995 *Fractal Image Compression*, Springer-Verlag, New York, New York, USA.
- [162] Fix E & Hodges JL 1951 'Discriminatory analysis- nonparametric discrimination: consistency properties', Report no.4, US Air Force School of Aviation Medicine, Random Field, Texas, USA.
- [163] Fong H 1996 'Pattern recognition in gray-level images by Fourier analysis', *Pattern Recognition Letters*, **17**(4):1477-1489.
- [164] Fox J 1990 'Describing univariate distributions'. In *Modern Methods of Data Analysis*, Fox J & Zong JS (eds), 53-125, Sage Publications, Newbury Park, California, USA.
- [165] , Fox J & Zong JS (eds) 1990 *Modern Methods of Data Analysis*, Sage Publications, Newbury Park, California, USA.
- [166] Freedman D & Diaconis P 1981 'On the histogram as a density estimator: L_2 theory', *Probability Theory and Related Fields*, **57**(4):453-476.
- [167] Freeman H 1974 'Computer Processing of Line-Drawing Images', *ACM Computing Surveys (CSUR)*, **6**(1):57-97.
- [168] Friedman JH 1989 'Regularised discriminant analysis', *Journal of the American Statistical Association*, **84**:165-175.
- [169] Frieman O, Farnebäck G & Werten C-F 2006 'A Bayesian approach for stochastic white matter tractography', *IEEE Transactions on Medical Imaging*, **25**(8): 965-978.
- [170] Fundana K, Overgaard NC & Heyden A 2008 'Variational segmentation of image segmentation of image sequences using region-based active contours and deformable shape priors', *International Journal of Computer Vision*, **80**(3):289-299.
- [171] Gardenier PH, McCallum BC & Bates RHT 1986 'Fourier transform magnitudes are unique pattern recognition templates', *Biological Cybernetics*, **54**(6):385-391.

- [172] Geman S & Geman D 1984 'Stochastic relaxation, Gibbs's distributions, and the Bayesian restoration of images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6**:721-741.
- [173] Geman S & Geman D 1987 'Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images', *Readings in Computer Vision: Issues, Problems, Principles, and Paradigms*, 564-584.
- [174] Genovese CR 2000 'A Bayesian tissue-course model for functional magnetic resonance imaging data', *Journal of the American Statistical Association*, **95**(451):691-719.
- [175] Ghosh A & Pal SK 1992 'Fuzzy geometry in image analysis', *Fuzzy Sets and Systems*, **48**(1):23-40.
- [176] Gikis HR, Richardson S & Spiegelhalter DJ (eds) 1995 *Markov Chain Monte Carlo In Practice*, Chapman & Hall, London, United Kingdom.
- [177] Gill PE, Murray D & Wright MH 1991 *Numerical Linear Algebra and Optimisation*, Addison-Wesely, Redwood City, California, USA.
- [178] Gnedenko BV 1943 Sur la distribution limite du terme maximum d'une série aléatoire *Annals of Mathematics*, **44**:423-453.
- [179] Goetcherian V 1980 'From binary to grey tone image processing using fuzzy logic concepts', *Pattern Recognition*, **12**(1):7-15.
- [180] Gooya A, Liao H, Matsumiya K, Masumiy K, Masamune K, Masutani Y & Dohi T 2008 'A variational method for geometric regularisation of vascular segmentation in medical images', *IEEE Transactions on Image Segmentation*, **17**(8):1295-1312.
- [181] Gong P 1993 'Change detection using principal component analysis and fuzzy set theory', *Canadian Journal of Remote Sensing*, **19**(1):22-22.
- [182] Gonzalez RC & Woods RE 2007 *Digital Image Processing*, (3rd edn), Pearson/Prentice Hall, Harlow, UK.
- [183] Goutsias J & Heijmans HJ AM 'Fundamenta Morphologicae Mathematicae', in Goutsias J & Heijmans HJ AM (eds) 2000 *Mathematical Morphology*, IOS Press, Amsterdam, The Netherlands.

- [184] Goutsias J & Heijmans HJ AM (eds) 2000 *Mathematical Morphology*, IOS Press, Amsterdam, The Netherlands.
- [185] Goutsias J, Vincent L & Bloomberg DS (eds) 2000 *Mathematical Morphology and its Applications to Image and Signal Processing*, Kluwer Academic Publishers, Norwell, Massachusetts, USA.
- [186] Green PJ 1990 'Bayesian reconstructions from emission tomography data using a modified EM algorithm', *IEEE Transactions in Medical Imaging*, **9**(1): 84-93.
- [187] Green P 'Markov Chain Monte Carlo in image analysis' in Gikis HR, Richardson S & Spiegelhalter DJ (eds) 1995 *Markov Chain Monte Carlo In Practice*, Chapman & Hall, London, United Kingdom.
- [188] Green PJ, Hjart NL & Richardson S 2003 *Highly Structured Stochastic Systems*, Oxford University Press, Norfolk, United Kingdom.
- [189] Grenander U & Miller MI 1994 'Representations of knowledge in complex systems', *Journal of the Royal Statistical Society Series B*, **56**(4):549-603.
- [190] Griffiths JA, Speller RD, Royle GJ, Horrocks A, Olivo A, Pani S, Longo R, Dreossi D, Spencer SH, Robbins MS, Clifford DP & Hanby AM 2003 'X-ray diffraction CT on excised breast tissue sections: first results from Elettra', *2003 IEEE Nuclear Science Symposium Conference Record*, **4**: 2330-2334.
- [191] Griffiths DF & Watson GA 1999 *Numerical Analysis*, Chapman & Hall/CRC, Boca Raton, Florida, USA.
- [192] Guliato D, Rangayyan RM, Carnielli WA, Zuffo JA & Desautels JEL 2003 'Segmentation of breast tumours in mammograms using fuzzy sets', *Journal of Electronic Imaging*, **12**:369-378.
- [193] Gumbel EJ 1958 *Statistics of Extremes*, Columbia University Press, New York, New York, USA.
- [194] Haar A 1910 'Zur Theorie der orthogonalen Funktionen-Systeme', *Mathematische Annalen*, **69**:331:371.

- [195] Hall LO, Bensaid M, Clarke LP & Velthuizen RP 1992 'A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain', *IEEE Transactions on Neural Networks*, **3**(5):672-682.
- [196] Hall P & Murison R 1993 'Correcting the negativity of high-order kernel density estimators', *Journal of Multivariate Analysis*, **47**(1):103-122.
- [197] Hall P & Patil P 1995a 'Formulae for mean-integrated squared error of nonlinear wavelet-based density estimators', *The Annals of Statistics*, **23**(3):905-928.
- [198] Hall P & Patil P 1995b 'On wavelet methods for estimating smooth functions', *Bernoulli*, **1**:41-58.
- [199] Hall P & Patil P 1996 'Effect of threshold rules on performance of wavelet-based curve estimators', *Statistica Sinica*, **6**:331-345.
- [200] Hall P & Wood A 1993 'On the performance of box-counting estimation of fractal dimension', *Biometrika*, **80**(1): 246-252.
- [201] Halsey TC, Jensen MH, Kadanoff LP, Procaccia I & Shraiman BI 1986 'Fractal measures and their singularities: the characterisation of strange sets', *Physics Review A*, **33**:1141-1151.
- [202] Hand DJ 1982 *Kernel Discriminant Analysis*, Research Studies Press, Chichester, United Kingdom.
- [203] Hanson KM & Wecksung GW 1983 'Bayesian approach to limited-angle reconstruction in computed tomography', *Journal of the Optical Society of America*, **73**(11):1501-1509.
- [204] Hanson KM 'Flexible prior models in Bayesian image analysis' in Mohammad-Djafari A & Demoment G (eds) 1993 *Maximum Entropy and Bayesian Methods*, Kluwer Academic, Dordrecht, The Netherlands.
- [205] Hanson KM 1993 'Introduction to Bayesian image analysis', *Proceedings of the SPIE*, **1898**:716-731.
- [206] Harding G, Kosanetzky J & Neitzel U 1987 'X-ray diffraction computed tomography', *Medical Physics*, **14**: 515-525.

- [207] Harding G & Schreiber B 1999 'Coherent x-ray scatter imaging and its applications in biomedical science and industry', *Radiation Physics & Chemistry*, **56**(1-2): 229-245.
- [208] Härdle W 1991 *Smoothing Techniques: with implementation in S*, Springer-Verlag, New York, New York, USA.
- [209] Hassanien AE, Ali JM & Nobuhara H 2004 'Detection of spiculated masses in mammograms based on fuzzy image processing', *Lecture Notes in Computer Science*, 1002-1007.
- [210] Hastie T, Buja A & Tibshirani R 1995 'Penalized discriminant analysis', *Annals of Statistics*, **23**(1):73-102.
- [211] Hastie T & Tibshirani R 1996 'Discriminant Adaptive Nearest Neighbor Classification', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(6):607-616.
- [212] Hastie T, Tibshirani R & Friedman J 2001 *The Elements of Statistical Learning*, Springer, New York, New York, USA.
- [213] Haykin S 1999 *Neural Networks: A comprehensive foundation* (2nd edn), Prentice-Hall, Englewood Cliffs, New Jersey, USA.
- [214] Heijmans HJAM 1995 'Mathematical morphology: a modern approach in image processing based on algebra and geometry', *SIAM Review*, **37**(1):1-36.
- [215] Heijmans HJAM & Roerdink JBTM (eds) 1998 *Mathematical Morphology and its Applications to Image and Signal Processing*, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- [216] Hermosillo G & Faugeras O 2001 'Dense image matching with global and local statistical criteria: a variational approach', *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **11**-73-I-78.
- [217] Hermosillo G, Ched d'Hotel C & Faugeras O 2004 'Variational methods for multimodal image matching', *International Journal of Computer Vision*, **50**(3):329-343.
- [218] Hertz J, Krogh A & Palmer RG 1991 *Introduction to the Theory of Neural Computation*, Addison Wesley, Reading, Massachusetts, USA.

- [219] Hewer GA, Kenny C & Manjunath BS 1998 'Variational image segmentation using boundary functions', *IEEE Transactions on Image Processing*, **7**(9):1269-1282.
- [220] Höppner F, Klawonn F, Kruse R, Runkler T 1999 *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, John Wiley and Sons, New York, New York, USA.
- [221] Horn M, Husby O & Rue H 'Advances in Bayesian Image Analysis' in Green PJ, Hjart NL & Richardson S 2003 *Highly Structured Stochastic Systems*, Oxford University Press, Norfolk, United Kingdom.
- [222] Hosking JRM 1985 'Algorithm AS 215: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution', *Journal of the Royal Statistical Society Series C*, **34**(3):301-310.
- [223] Howell A, Grossman JG, Cheung KANC, Kanbi L, Evans DGR & Hasnain SS 2000 'Can hair be used to diagnose breast cancer?', *Journal of Medical Genetics*, **37**(4): 297-298.
- [224] Huang LK & Wang MJJ 1995 'Image thresholding by minimising measures of fuzziness', *Pattern Recognition*, **28**(1):41-51.
- [225] Hunt F 1990 'Error analysis and convergence of capacity dimension algorithms', *SIAM Journal of Applied Mathematics*, **50**: 307-321.
- [226] Huntsberger TL, Jacobs CL & Cannon RL 1985 'Iterative fuzzy image segmentation', *Pattern Recognition*, **18**(2):131-138.
- [227] IEEE Standards Committee 754, *IEEE Standard for Binary Floating-Point Arithmetic, ANSI/IEEE Standard 754-1985*, Institute of Electrical and Electronics Engineers, New York, 1985. Reprinted in SIGPLAN Notices, **22**(2):925, 1987.
- [228] Ivanov PC, Amaral LAN, Goldberger AL, Haulin S, Rosenblum MG, Struzik ZR & Stanley HE 1999 'Multifractality in human heart beat dynamics', *Nature*, **6735**:461-465.
- [229] Iwanowski M & Serra J 'The morphological-affine object deformation' in Goutsias J, Vincent L & Bloomberg DS (eds) 2000 *Mathematical Morphology and its Applications to Image and Signal Processing*, Kluwer Academic Publishers, Norwell, Massachusetts, USA.

- [230] Jähne B 2005 *Digital Image Processing* (6th edn), Springer, Berlin, Germany.
- [231] Jahns G, Nielsen HM & Paul W 2001 'Measuring image analysis attributes and modelling fuzzy consumer aspects of tomato quality grading', *Computers and Electronics in Agriculture*, **31**(1):17-29.
- [232] Jain A & Jain J 1978 'Partial differential equations and finite difference methods in image processing - Part II image restoration', *IEEE Transactions on Automatic Control*, **23**(5):817-834.
- [233] Jain AK, Zhong Y & Lakshmanan S 1996 'Object matching using deformable templates', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **18**(3):267-278.
- [234] Jain AK, Zhang Y & Dubuisson-Jolly MP 1998 'Deformable template models: a review', *Signal Processing*, **71**(2):109-129.
- [235] James V, Kearsley J, Irving T, Amemiya Y & Cookson D 1999 'Using hair to screen for breast cancer', *Nature*, **398**: 33-34.
- [236] James VJ 2001 'The importance of good images in using hair to screen for breast cancer', *Journal of Medical Genetics*, **38**: e16.
- [237] James VJ 2003 'The traps and pitfalls inherent in the correlation of changes in the fibre diffraction pattern of hair with breast cancer', *Physics in Medicine and Biology*, **48**: L5-L9.
- [238] James VJ 2004 'Using hair to screen for breast cancer', US Patent 6,718,007.
- [239] James V, Corino G, Robertson T, Dutton N, Halas D, Boyd A, Bentel J & Papadimitriou J 2005 'Early diagnosis of breast cancer by hair diffraction', *International Journal of Cancer*, **114**(6): 969-972.
- [240] James VJ 2006 'A place for fiber diffraction in the detection of breast cancer?' *Cancer Detection & Prevention*, **30**(3): 233-238.
- [241] Jang BK 1990 *Shape Analysis Using Mathematical Morphology*, Doctoral Thesis, The University of Wisconsin-Madison, USA.

- [242] Javidi B 1990 'Generalization of the linear matched filter concept to nonlinear matched filters', *Applied Optics*, **29**(8):1215-1224.
- [243] Javidi B & Horner JL 1998 'Pattern recognition with nonlinear techniques in the Fourier domain', *Progress in Optics*, **38**:345-418.
- [244] Jeffrey A 2000 *Handbook of Mathematical Formulas & Integrals* (2nd edn), Academic Press, London, United Kingdom.
- [245] Jenkinson AF 1955 'The frequency distribution of the annual maximum (or minimum) of meteorological elements', *Quarterly Journal of the Royal Meteorological Society*, **81**:158-171.
- [246] Jerri AJ 1977 'The Shannon sampling theorem - its various extensions and applications: A tutorial review', *Proceedings of the IEEE*, **65**(1): 1565-1595.
- [247] John RI, Innocent PR & Barnes MR 2000 'Neuro-fuzzy clustering of radiographic tibia image data using type 2 fuzzy sets', *Information Sciences*, **125**(1-4):65-82.
- [248] Johns PC & Yaffe MJ 1983 'Coherent scatter in diagnostic radiology', *Medical Physics*, **10**(1): 40-50.
- [249] Kadler KE, Holmes DF, Trotter JA & Chapman JA 1996 'Collagen fibril formation', *Biochemistry Journal*, **316**:1-11.
- [250] Kalifa J, Mallat S & Rouge B 1998 'Image deconvolution in mirror wavelet bases', *ICIP98, Proceedings 1998 International Conference on Image Processing*, **1**:565-569.
- [251] Kalifa J, Mallat S & Rouge B 2003 'Deconvolution by thresholding in mirror wavelet bases', *IEEE Transactions on Image Processing*, **12**(4): 446-457.
- [252] Kandel A 1982 *Fuzzy Techniques in Pattern Recognition*, Wiley-Interscience, New York, New York, USA.
- [253] Kass M, Witkin A & Terzopoulos D 1988 'Snakes: Active Contour Models', *International Journal of Computer Vision*, **1**:321-331.
- [254] Katartzis A, Sahli H, Cornelis J, Costaridou L & Panayiotakis G 'A MRF-based approach for the measurement of skin thickness' in Costaridou L (ed) 2005 *Medical Image Analysis Methods*, Taylor & Francis, Boca Raton, Florida, USA.

- [255] Kauppila, S, Stenbäck, F, Risteli, J, Jukkola, A & Risteli, L 1998 'Aberrant type I and type III collagen gene expression in breast cancer in vivo', *Journal of Pathology*, **186**: 262-268.
- [256] Kendall DG, Barden D, Carne TK & Le H 1999 *Shape and Shape Theory*, Wiley, New York, New York, USA.
- [257] Kerre EE & Nachtegaal M (eds) 2000 *Fuzzy Techniques in Image Processing*, Physica-Verlag, New York, New York, USA.
- [258] Kidane G, Speller RD, Royle GJ & Hanby AM 1999 'X-ray scatter signatures for normal and neoplastic breast tissue', *Physics in Medicine and Biology*, **44**: 1791-1802.
- [259] Kimmel R & Sochen NA 1999 'Geometric variational approach for color image enhancement and segmentation' in Nielsen M, Johansen P, Olsen OF & Weickert J 1999 *Scale-space Theories in Computer Vision*, Springer-Verlag, Berlin Heidelberg, Germany.
- [260] Kingsbury NG 1999 'Image processing with complex wavelets', *Philosophical Transactions-Royal Society of London Series A: Mathematical, Physical and Engineering Sciences*, **1760**: 2543-2560.
- [261] Klir GJ & Yuan B 1995 *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, New Jersey, USA.
- [262] Koepfler G, Lopez C & Morel J-M 1994 'A multi-scale algorithm for image segmentation by variational method', *SIAM Journal of Numerical Analysis*, **31**(1):282-299.
- [263] Kornprobst P, Deriche R & Aubert G 1999 'Image sequence analysis via partial differential equations', *Journal of Mathematical Imaging and Vision*, **11**(1):5-26.
- [264] Kosanetzky J, Knoerr B, Harding G & Nietzel U 1987 'X-ray diffraction measurements of some plastic materials and body tissues', *Medical Physics*, **14**: 526-532.
- [265] Kotel'nikov VA 1933 'On the transmission capacity of 'ether' and wire in electrocommunications', *Izd. Red. Upr. Svyaz: RKKA*, Moscow, USSR.
- [266] Kotz S & Nadarajah S 2000 *Extreme Value Distributions: Theory and Applications*, Imperial College Press, London, UK.

- [267] Kuhl EP & Giardina CR 1982 'Elliptical Fourier features of a closed contour', *Computer Graphics and Image Processing*, **18**(3):236-258.
- [268] Kundur D & Hatzinakos D 1998 'Digital watermarking using multiresolution wavelet decomposition', *IEEE International Conference on Acoustics, Speech and Signal Processing*, 5fV-2969-V-2972.
- [269] Kuo YH, Lee C & Liu CC 1997 'A new fuzzy edge detection method for image enhancement', *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems*, **2**:1069-1074.
- [270] Laaziri K, Sutton M, Ghadirian P, Scott AS, Paradis A-J, Tonin PN & Foulkes WD 2002 'Is there a correlation between the structure of hair and breast cancer or BRCA1/2 mutations ?' *Physics in Medicine and Biology*, **47**: 1623-1632.
- [271] Lachenbruch PA & Goldstein M 1979 'Discriminant analysis', *Biometrics*, **5**(3):9-85.
- [272] Larrey-Ruiz K 2008 'A Fourier domain framework for variational image registration', *Journal of Mathematical Imaging & Vision*, **32**(1):57-72.
- [273] Lazarev P, Paukshto M , Pelc N & Sakharova A 2000 'Human tissue x-ray diffraction: breast, brain and prostate', *Proceedings of the 22nd Annual International Conference in the IEEE Engineering in Medicine & Biology Society*, **4**: 3230-3233.
- [274] Leclair RJ & Johns PC 1998 ' A semianalytic model to investigate the potential applications of x-ray scatter imaging', *Medical Physics*, **35**(6): 1008-1020.
- [275] Leclair RJ & Johns PC 1999 'Analysis of spectral blur effects in x-ray scatter imaging', *Medical Physics*, **26**(9): 1811-1816.
- [276] Leclair RJ & Johns PC 2001 'X-ray forward scatter imaging: Experimental validation of model', *Medical Physics*, **28**(2): 210-219.
- [277] Lee SJ, Rangarajan A & Gindi G 1995 'Bayesian image reconstruction in SPECT using high order mechanical models as priors', *IEEE Transactions on Medical Imaging*, **14**(4):669-680.
- [278] Lee EK, Cook D, Klinke S & Lumley T 2005 ' Projection pursuit for exploratory supervised classification', *Journal of Computational and Graphical Statistics*, **14**(4): 831-846.

- [279] Le Flanchec V, Gazeau D, Taboury J & Zemb T 1996 'Two-dimensional desmearing of centrosymmetric small-angle x-ray scattering diffraction patterns', *Journal of Applied Crystallography*, **29**(2):110-117.
- [280] Lele SR & Richtsmeir JT 2000 *An Invariant Approach to the Statistical Analysis of Shapes*, Chapman & Hall/CRC, Boca Raton, Florida, USA.
- [281] Lelieveldt B, Sonka M, Bolinger L, Scholz TD, Kayser H, van der Geest R & Reiber JHC 2000 'Anatomical modeling with fuzzy implicit surface templates: applications to automated localization of the heart and lungs in thoracic MR volumes', *Computer Vision and Understanding*, **1**:1-20.
- [282] Lestrel P 'Introduction and overview of Fourier descriptors' in Lestrel P 1997 *Fourier Descriptors and their Applications in Biology*, Cambridge University Press, New York, New York, USA.
- [283] Lestrel P 1997 *Fourier Descriptors and their Applications in Biology*, Cambridge University Press, New York, New York, USA.
- [284] Leung SH, Wang SL & Lau WH 2004 'Lip image segmentation using fuzzy clustering incorporating an elliptic shape function', *IEEE Transactions on Image Processing*, **13**(1):51-62.
- [285] Lewis AS & Knowles G 1992 'Image compression using the 2-D wavelet transform', *IEEE Transactions on Image Processing*, **1**(2): 244-250.
- [286] Lewis RA 1994 'Multiwire gas proportional counters: decrepit antiques or classic performers?' *Journal of Synchrotron Radiation*, **1**:43-53.
- [287] Lewis RA, Hall CJ, Parker B, Jones A, Helsby W, Sheldon JW, Clifford P, Hillon M & Fore N 1997 'The Daresbury fast multi-wire linear X-ray detector system, *Nuclear Instruments and Methods in Physics Research Section A*, **392**(1-3):42-46.
- [288] Lewis RA, Helsby WI, Jones AO, Hall CJ, Parker B, Sheldon J, Clifford P, Hillen M, Sumner I, Fore NS, Jones RWM & Roberts KM 1997 'The 'RAPID' high rate large area X-ray detector system', *Nuclear Instruments and Methods in Physics Research Section A*, **392**(1-3):32-41.

- [289] Lewis RA, Rogers KD, Hall CJ, Towns-Andrews E, Slawson S, Evans A, Pinder SE, Ellis E, Boggis IO & Hufton CRM 2000 'Breast cancer diagnosis using scattered X-rays', *Journal of Synchrotron Radiation*, **7**(5):348-352.
- [290] Li S & Li W 2000 'Shape-adaptive discrete wavelet transforms for arbitrarily shaped visual object coding', *IEEE Transactions on Circuits and Systems for Video Technology*, **10**(5):725-743.
- [291] Li M & Yuan B 2005 '2D-LDA: A statistical linear discriminant analysis for image matrix', *Pattern Recognition Letters*, **26**(5):527-532.
- [292] Liang LR & Looney CG 2003 'Competitive fuzzy edge detection', *Applied Soft Computing*, **3**(2):123-137.
- [293] Liew AWC, Leung SH & Lau WH 2000 'Fuzzy image clustering incorporating spatial continuity', *IEEE Proceedings, Vision, Image and Signal Processing*, **147**(2):185-192.
- [294] Likas AC & Galatsanos NP 2004 'A variational approach to blind image deconvolution', *IEEE Transactions on Signal Processing*, **52**(8):2222-2233.
- [295] Lim YW & Lee SU 1989 'On the color image segmentation algorithm based on the thresholding and the fuzzy c-means techniques', *Pattern Recognition*, **23**(9):935-952.
- [296] Lin J 1991 'Divergence measures based on the Shannon entropy', *IEEE Transactions on Information Theory*, **37**(1):145-151.
- [297] Lin JS, Cheng KS & Mao CW 1996 'A fuzzy Hopfield neural network for medical image segmentation', *IEEE Transactions on Nuclear Science*, **43**(4:Part 2):2389-2398.
- [298] Loader C 1999 *Local regression and likelihood*, Springer, New York, New York, USA.
- [299] Lovejoy S & Schertzer D 1989 'Multifractals, universality classes and satellite and radar measurements of cloud and rain fields', *Journal of Geophysical Research-Atmosphere*, **95**(D3):2021-2034.
- [300] Lundahl T, Ohley WJ, Kuklinski WS, Williams DO, Gerwitz H, Most AS 1985 'Analysis and interpolation of angiographic images by the use of fractals', *IEEE Computers in Cardiology*: 355-358.

- [301] Lysakes ML & Tai AXC 2003 'Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time', *IEEE Transactions in Image Processing*, **12**(12):1579-1590.
- [302] Mallat S 1998 *A Wavelet Tour of Signal Processing*, (2nd edn), Academic Press, San Diego, California, USA.
- [303] Mallat S & Hwang WL 1992 'Singularity detection and processing with wavelets', *IEEE Transactions on Information Theory*, **38**(2):617-643.
- [304] Mallat S & Zhang Z 1992 'Characterization of signals from multiscale edges', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(7): 710-732.
- [305] Mallat S & Zhang Z 1993 'Matching Pursuit with Time-Frequency Dictionaries', *IEEE Transactions on Signal Processing*, **41**(12): 3397-3415.
- [306] Mandelbrot B 1982 *The Fractal Geometry of Nature*, WH Freeman, New York, New York, USA.
- [307] Mandelbrot B 2006 *The Fractal Geometry of Nature*, (21st edn), WH Freeman, San Francisco, California, USA.
- [308] Manousaki AG, Manios AG, Tsompanaki EI & Tosca AD 2006 'Use of color texture in determining the nature of melanocytic skin lesions-a qualitative and quantitative approach', *Computers in Biology and Medicine*, **36**:419-427.
- [309] Mansfield JR, Sawa MG, Payette JR, Abdulrauf B, Stranc MF & Mantsch HH 1998 'Tissue viability by multispectral near infrared imaging: A fuzzy c-means clustering analysis', *IEEE Transactions on Medical Imaging*, **17**(6):1011-1018.
- [310] Mansoor AB, Mian AS, Khan A & Khan SA 2007 'Fuzzy morphology for edge detection and segmentation', *Lecture Notes in Computer Science*, **4842**:811-821.
- [311] Mansouri AR 2002 'Region tracking via level set PDEs without motion computation', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24**(7):947-961.
- [312] March R & Dozio M 1997 'A variational method for the recovery of smooth boundaries', *Image and Vision Computing*, **15**(9):705-712.

- [313] Marron JS & Wand MP 1992 'Exact mean integrated squared error', *The Annals of Statistics*, **20**(2):712-736.
- [314] Masry E 1994 'Probability density estimation from dependent observations using wavelet orthonormal bases', *Statistics & Probability Letters*, **21**(3):181-194.
- [315] Matheron G 1975 *Random sets and Integral Geometry*, John Wiley and Sons, New York, New York, USA.
- [316] Matsumoto ML & Nishssimura T 1998 'Mersenne Twister: A 623-dimensionality equidistributed uniform pseudorandom number generator', *ACM Transactions on Modeling and Computer Simulation*, **31**(10):1192-1201.
- [317] McInerney T & Terzopoulos D 1988 'Deformable models in medical image analysis: a survey', *Medical Image Analysis*, **1**(2):91-108.
- [318] Meneveau C & Sreenivasen KR 2006 'The multifractal nature of turbulent energy dissipation', *Journal of Fluid Mechanics Digital Archive*, **224**:429-484.
- [319] Meyer P, Goergl R, Botz JW & Fratzl P 2000 'Breast cancer screening using small-angle x-ray scattering analysis of human hair', *Journal of the National Cancer Institute*, **92**(13): 1092-1093.
- [320] Meyer P & James VJ 2001 'Experimental confirmation of a distinctive diffraction pattern in hair from women with breast cancer', *Journal of the National Cancer Institute*, **93**(1): 873-875.
- [321] Mihcak MK, Kozintsev I, Ramchandran K & Moulin P 1999 'Low complexity image denoising based on statistical modeling of wavelet coefficients', *IEEE Signal Processing Letters*, **6**(12):300-303.
- [322] Mix DF & Olejniczak KJ 2003 *Elements of Wavelets for Engineers and Scientists*, John Wiley & Sons, Hoboken, New Jersey, USA.
- [323] Moghaddamzadeh A & Bourbakis N 1997 'A fuzzy region growing approach for segmentation of color images', *Pattern Recognition*, **30**(6):867-881.
- [324] Mohammad-Djafari A & Demoment G (eds) 1993 *Maximum Entropy and Bayesian Methods*, Kluwer Academic, Dordrecht, The Netherlands.

- [325] Molina R, Mateos J & Katsaggelos AK 2006 'Blind image deconvolution using a variational approach for parameter, image, and blur estimation', *IEEE Transactions on Image Processing*, **15**(12):3715-3727.
- [326] Morel J-M & Solimini S 1995, 'Variational methods in image segmentation', *Progress in Nonlinear Differential Equations and their Applications*, **14**:245, Birkhäuser, Boston, Massachusetts, USA.
- [327] Morgan JN & Sonquist JA 1963 'Problems in the analysis of survey data, and a proposal', *Journal of the American Statistical Association*, **58**:415-434.
- [328] Morgan JN & Messenger RC 1973 *THAID: A Sequential Search Program for the Analysis of Nominal Scale Dependent Variables*, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA.
- [329] Morillas S, Schulte S, Kerre EE & Peris-Fajarnes G 2007 'A new fuzzy impulse noise detection method for color images', *Lecture Notes in Computer Science*, **4522**:492-501.
- [330] Morse SP 1968 'A mathematical model for the analysis of contour-line data', *Journal of the ACM (JACM)*, **15**(2):205-220.
- [331] Mueller P & Vidakovic B 1998 'Bayesian inference with wavelets: density estimation', *Journal of Computational & Graphical Statistics*, **7**(4):456-468.
- [332] Mumcuoglu EU, Leahy R & Cherry SR 1994 'Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images', *IEEE Transactions on Medical Imaging*, **13**(4): 687-701.
- [333] Mumford D & Shah J 1989 'Optimal approximations by piecewise smooth functions and associated variational problems', *Communications on Pure & Applied Mathematics*, **XLII**:577-685.
- [334] Muntz EP, Fewell T, Jennings R & Bernstein H 1983 'On the significance of very small angle scattered radiation to radiographic imaging at low energies', *Medical Physics*, **10**(6): 819-823.
- [335] Murshed SSN & Kropatsch W (eds) 2001, *ICAPR 2001, LNCS 2013*, Springer-Verlag, Berlin Heidelberg, Germany.

- [336] Muzy JF, Bacry E & Arnéodo A 1991 'Wavelets and multifractal formalism for singular signals: application to turbulence data', *Physical Review Letters*, **67**(25): 3515-3518.
- [337] Nachtegaal M, Van der Weken D, Van de Ville D & Kerre EE 2003 *Fuzzy Filters for Image Processing*, Springer-Verlag, New York, New York, USA.
- [338] Nair MS & Wilsey M 2008 'Modified method for denoising color images using fuzzy approach', *Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel, Distributed Computing 2008*, pp. 507-512.
- [339] Neelamani R, Choi H & Baraniuk R 2004 'ForWaRD: Fourier-wavelet regularised deconvolution for ill-conditioned systems', *IEEE Transactions on Image Processing*, **52**(3): 418-433.
- [340] Neter J, Kutner MH, Nachtsheim CJ & Wasserman W 1996 *Applied Linear Statistical Models*, McGraw-Hill, Boston, Massachusetts, USA.
- [341] Nguyen HH & Cohen P 1993 'Gibbs random fields fuzzy clustering, and the unsupervised segmentation of textured images', *CVGIP Graphical Models and Image Processing*, **55**(1):1-19.
- [342] Nielsen M, Johansen P, Olsen OF & Weickert J 1999 *Scale-space Theories in Computer Vision*, Springer-Verlag, Berlin Heidelberg, Germany.
- [343] Nikolova M 2004 'A variational approach to remove outliers and impulse noise', *Journal of Mathematical Imaging and Vision*, **20**(1/2):99-120.
- [344] Nixon MS & Aguado AS 2002 *Feature Extraction and Image Processing*, Newnes publications, Jordan Hill, Oxford, United Kingdom.
- [345] Nordstrom KN 1990 'Biased anisotropic diffusion: A unified regularisation and diffusion approach to edge detection', *Image Vision and Computing*, **8**(4):318-327.
- [346] Nyquist H 1928 'Certain topics in telegraph transmission theory', *AIEE Transactions*, **47**: 617-644 in 'Classic paper' 2002, *Proceedings of the IEEE*, **90**(2): 280-305.
- [347] Ooi G, Fox J, Siu K, Lewis R, Bambery KR, McNaughton D & Wood BR 2008 'Fourier transform infrared imaging and small angle x-ray scattering as a combined biomolecular approach to diagnosis of breast cancer', *Medical Physics*, **35**(5): 2151-2161.

- [348] Oren M, Papageorgiou C, Sinha P, Osuna E & Poggio T 1997 'Pedestrian detection using wavelet templates', *Proceedings 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.193-199.
- [349] Pal SK & King RA 1980 'Image enhancement using fuzzy set', *Electronics Letters*, **16**(10):376-378.
- [350] Pal SK & Dutta-Majumder DK 1986 *Fuzzy Mathematical Approach to Pattern Recognition*, Halsted Press, New York, New York, USA.
- [351] Pal SK 1992 'Fuzzy sets in image processing and recognition', *IEEE International Conference on Fuzzy Systems*, 119-126.
- [352] Pal SK & Mitra S 1999 *Neuro-fuzzy Pattern Recognition: Methods in Soft Computing*, John Wiley and Sons, New York, New York, USA.
- [353] Pandey G, Lovejoy S & Schertzer D 1998 'Multifractal analysis of daily river flows including extremes for basins of five to two million square kilometres, one day to 75 years', *Journal of Hydrology*, **208**(1-2): 62-81.
- [354] Papodemetris X, Sinusas AJ, Dione DP & Duncan JS 2001 'Estimation of 3D left ventricular deformation from echocardiography', *Medical Image Analysis*, **5**(1):17-28.
- [355] Paragios N & Deriche R 2000 'Geodesic active contours and level sets for the detection of the detection and tracking of moving objects', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **23**(3):266-280.
- [356] Park W, Hoffman EA & Sonka M 1998 'Segmentation of intrathoracic airway trees: a fuzzy logic approach', *IEEE Transactions on Medical Imaging*, **17**(4):489-497.
- [357] Parzen E 1962 'On estimation of a probability density function and mode', *The Annals of Mathematical Statistics*, **33**(3):1065-1076.
- [358] Pavlopoulos S, Kyriacou E, Koutsouris D, Blekas K, Stafylopatis A & Zoumpoulis P 2000 'Fuzzy neural network-based textural analysis of ultrasonic images', *IEEE Engineering in Medicine and Biology Magazine*, **19**(1):39-47.

- [359] Pearson SJ, Siu KKW, Hall C, Reid C & Falzon G 2006 'Small-angle x-ray scattering and second-harmonic generation imaging studies of collagen in invasive carcinoma', *RiverPhys: Australian Institute of Physics 17th National Congress 2006*, Brisbane, Queensland, Australia, retrieved 12 November 2008 from <http://www.aip.org.au/content/congress2006#bmp>.
- [360] Pedrycz W 1990 'Fuzzy sets in pattern recognition: methodology and methods', *Pattern Recognition*, **23**(1-2):121-146.
- [361] Peirce BO 1899 *A Short Table of Integrals*, Ginn and Co, Boston, Massachusetts, USA.
- [362] Penev S & Dechevsky L 1997 'On non-negative wavelet-based density estimators', *Journal of Nonparametric Statistics*, **7**:365-394.
- [363] Peplow DE & Verghese K 1998 'Measured molecular coherent scattering form factors of animal tissues, plastic and human breast tissues', *Physics in Medicine and Biology*, **43**: 2431-2452.
- [364] Perez P, Gangnet M & Blake A 2003 'Poisson image editing', *ACM Transactions on Graphics*, **22**(3):313-318.
- [365] Persoon E & Fu KS 1977 'Shape discrimination using Fourier descriptors', *IEEE Transactions on Systems, Man & Cybernetics*, **7**(3):170-179.
- [366] Perona P & Malik J 1990 'Scale-space and edge detection using anisotropic diffusion', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(7):629-639.
- [367] Peters II RA 2008 'EECE/CS 253 Image Processing' course notes, School of Engineering, Vanderbilt University from www.archive.org/details/Lectures_on_Image_Processing, updated 28/4/2008, accessed 8/2/2009.
- [368] Petigen H-O, Jürgens H & Saupe H 1992, *Chaos & Fractals: new frontiers of science*, Springer-Verlag, New York, New York, USA.
- [369] Petrosino A & Salvi G 2006 'Rough fuzzy sets based scale space transforms and their use in image analysis', *International Journal of Approximate Reasoning*, **41**(2):212-228.

- [370] Phoong S-M, Kim CW, Vaidyanathan PP & Ansari R 1995 'A new class of two-channel biorthogonal filter banks and wavelet bases', *IEEE Transactions on Signal Processing*, **43**(3):649-665.
- [371] Pizer SM, Fritsch DS, Low K-C & Furst JD '2D and 3D figural models of anatomic objects from medical images' in Heijmans HJAM & Roerdink JBTM (eds) 1998 *Mathematical Morphology and its Applications to Image and Signal Processing*, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- [372] Po DD-Y & Do MN 2006 'Directional multiscale modeling of images using the contourlet transform', *IEEE Transactions on Image Processing*, **15**(6):1610-1620.
- [373] Pollak I, Willsky AS & Krim H 2000 'Image segmentation and edge enhancement with stabilised inverse diffusion equations', *IEEE Transactions on Image Processing*, **9**(21):256-266.
- [374] Poletti ME, Gonçalves OD & Mazzaro I 2002a 'X-ray scattering from human breast tissues and breast-equivalent materials', *Physics in Medicine and Biology*, **47**(1): 47-64.
- [375] Poletti ME, Gonçalves OD & Mazzaro I 2002b 'Coherent and incoherent scattering of 17.44 and 6.93 keV x-ray photons scattered from biological and biological-equivalent samples: characterisation of tissue', *X-ray Spectrometry*, **31**(1): 57-61.
- [376] Portilla J & Simoncelli EP 2000 'A parametric texture model based on joint statistics of complex wavelet coefficients', *International Journal of Computer Vision*, **40**(1):49-71.
- [377] Portilla J, Strela V, Wainwright MJ & Simoncelli EP 2003 'Image denoising using scale mixtures of gaussians in the wavelet domain', *IEEE Transactions on Image Processing*, **12**(11): 1338-1351.
- [378] Prasad L & Iyengar SS 1997 *Wavelet Analysis with Applications to Image Processing*, CRC Press, Boca Raton, Florida, USA.
- [379] Pucci-Minafra I, Luparello C, Andriolo M & Basirico L 1993 'A new form of tumour and fetal collagen that binds laminin', *Biochemistry*, **32**(29):7421-7427.
- [380] Rangayyan RM 2005 *Biomedical Image Analysis*, CRC Press, Boca Raton, Florida, USA.

- [381] Rao CR 1948 The utilization of multiple measurements in problems of biological classification (with discussion)', *Journal of the Royal Statistical Society Series B*, **10**:159-203.
- [382] Reid C 2006 *A Multi-Modality Imaging Study of the Supra-molecular Structural Changes of Collagen Induced by Breast Cancer*, Honours thesis, Department of Physics & Electronics, University of New England, NSW, Australia.
- [383] Rezza FM 1961 *An Introduction to Information Theory*, McGraw-Hill, New York, New York, USA.
- [384] Rezaee MR, van der Zwert PMJ, Lelieveldt BPF, van der Greest RJ & Reiber JHC 2000 'A multiresolution image segmentation technique based on pyramidal segmentation and fuzzy clustering', *IEEE Transactions on Image Processing*, **9**(7):1238-1248.
- [385] Ripley BD 1996 *Pattern Recognition and Neural Networks*, Cambridge University Press, New York, New York, USA.
- [386] Rogers K, Wilkinson S, Round A, Hall C 2006 'Cancer diagnosis using SAXS imaging', *Powder Diffraction*, **21**: 169.
- [387] Rosales-Silva A, Ponmaryov VI & Gallegos-Funes FJ 2007 'Fuzzy vector directional filters for multichannel image denoising', *Lecture Notes in Computer Science*, **4756**:124-133.
- [388] Rosenblatt M 1956 'Remarks on some nonparametric estimates of a density function', *The Annals of Mathematical Statistics*, **27**(3):832-837.
- [389] Rossi F & Villa N 2006 'Support vector machine for functional data classification', *Neurocomputing*, **69**(7-9):730-742.
- [390] Round AR, Wilkinson SJ, Hall CJ, Rogers KD, Glatter O, Wess T & Ellis IO 2005 'A preliminary study of breast cancer diagnosis using laboratory based small angle x-ray scattering', *Physics in Medicine and Biology*, **50**: 4159-4168.
- [391] Round AR 2006 '*Ultra-structural analysis of breast tissue*', Doctoral Thesis, Postgraduate Medical School, Department of Materials and Medical Sciences, Cranfield University, Cranfield, UK.

- [392] Royle GJ & Speller RD 1991 'Low-angle X-ray scattering for bone analysis', *Physics in Medicine and Biology*, **36**: 383-389.
- [393] Royle, GJ, Farquharson, M, Speller, R, Kidane, G, 1999 'Applications of X-ray diffraction analysis in crystalline and amorphous body tissues', *Radiation Physics and Chemistry*, **56**: 247-258.
- [394] Royston JP 1982 'An extension of Shapiro and Wilk's W test for normality to large samples', *Journal of the Royal Statistical Society Series C*, **31**(2):115-124.
- [395] Royston P 1992 'Approximating the Shapiro-Wilk W -test for non-normality', *Statistics & Computing*, **2**(3):117-119.
- [396] Ruanaidh O, Dowling JJK & Boland FM 1996 'Watermarking digital images for copyright protection', *IEEE Proceedings Vision, Image and Signal Processing*, **143**(4): 250-256.
- [397] Rudin LI, Osher S & Fatemi E 1992 'Nonlinear total variation based noise removal algorithms', *Physica D*, **60**:259-268.
- [398] Rudin LI & Osher S 1994 'Total variation based image restoration with free local constraints', *Proceedings ICIP-94. IEEE International Conference on Image Processing*, **1**:31-35.
- [399] Rue H & Held L 2005 *Gaussian Markov Random Fields: Theory and Applications*, Chapman & Hall/CRC, Boca Raton, Florida, USA.
- [400] Rumelhart DE & McClelland JL (eds) 1986 *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1. Foundations*, The MIT Press, Cambridge, Massachusetts, USA.
- [401] Ruppert D, Wand M & Carroll RJ 2003 *Semiparametric Regression*, Cambridge University Press, New York, New York, USA.
- [402] Russ JC 1992 *The Image Processing Handbook*, Boca Raton, Florida, USA.
- [403] Russo F 1998 'Edge detection in noisy images using fuzzy reasoning', *IEEE Instrumentation and Measurement Technology Conference, 1998. IMTC/98 Conference Proceedings*, **1**:369-372.

- [404] Ryan EA & Farquharson MJ 2004 'Angular dispersive x-ray scattering from breast tissue using synchrotron radiation', *Radiation Physics and Chemistry*, **71**(3-4): 971-972.
- [405] Ryan EA & Farquharson MJ 2007 'Breast tissue classification using x-ray scattering measurements and multivariate data analysis', *Physics in Medicine and Biology*, **52**: 6679-6696.
- [406] Samson C, Blanc-Feraud L, Aubert G & Zerubia J 2000 'A variational model for image classification and restoration', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22**(5):460-472.
- [407] Sapiro G 2001 *Geometric Partial Differential Equations in Image Analysis*, Cambridge University Press, New York, New York, USA.
- [408] Scarff R 1981 *Histological Typing of Breast Tumours*, (2nd edn), World Health Organisation, Geneva, Switzerland.
- [409] Schertzer D & Lovejoy S 1985 'Generalised scale invariance in turbulent phenomena', *Physics & Chemistry in Hydrology Journal*, **6**(623): 1233-1250.
- [410] Schlomka J-P, Schneider SM & Harding GL 2000 'Novel concept for coherent scatter x-ray computed tomography in medical applications', *Proceedings of the SPIE*, **4142**: 218-224.
- [411] Schulte S, Morilla S, Gregori V & Kerre EE 2007 'A new fuzzy color correlated impulse noise reduction method', *IEEE Transactions on Image Processing*, **16**(10):2565-2575.
- [412] Schwartz SC 1967 'Estimation of probability density by an orthogonal series', *The Annals of Mathematical Statistics*, **38**(4):1261-1265.
- [413] Scott DW 1979 'On optimal and data-based histograms', *Biometrika*, **66**(3):605-610.
- [414] Scott DW 1992 *Multivariate Density Estimation: Theory, Practice, and Visualization*, John Wiley & Sons, New York, New York, USA.
- [415] Serra J 1982 *Image Analysis and Mathematical Morphology*, Academic Press, London, United Kingdom.

- [416] Shannon CE 1948 'A mathematical theory of communication', *The Bell System Technical Journal*, **27**(623-656):379-423.
- [417] Shannon CE 1949 'Communications in the presence of noise', *Proceedings of the the IRE*, **37**: 10-21.
- [418] Shapiro SS & Wilk MB 1965 'An analysis of variance test for normality (complete samples)', *Biometrika*, **52**(3-4):591-611.
- [419] Sheather SJ & Jones MC 1991 'A reliable data-based bandwidth selection method for kernel density estimation', *Journal of the Royal Statistical Society Series B*, **53**(3):683-690.
- [420] Shen D & Ip HHS 1999 'Discriminative wavelet shape descriptors for recognition of 2D patterns', *Pattern Recognition*, **32**(2):151-166.
- [421] Sheng Y & Shen L 1994 'Orthogonal Fourier-Mellin moments for invariant pattern recognition', **11**(6):1748-1757.
- [422] Sherlock BG, Monro DM & Millard K 1994 'Fingerprint enhancement by directional Fourier filtering', *IEEE Proceedings, Vision, Image & Signal Processing*, **141**(2):87-94.
- [423] Sidhu S, Siu KKW, Falzon G, Nazaretian S, Hart SA, Fox JG, Susil BJ & Lewis RA 2008 'X-ray scattering for classifying tissue types with breast disease', *Medical Physics*, **35**(10): 4660-4670.
- [424] Silverman BW 1986 *Density Estimation for Statistics and Data Analysis*, Chapman & Hall/CRC, New York, New York, USA.
- [425] Simoncelli EP & Freeman WT 1995 'The steerable pyramid: a flexible architecture for multi-scale derivative computation', *International Conference on Image Processing*, **3**: 444-447.
- [426] Simoncelli EP 1999 'Modeling the joint statistics of images in the wavelet domain', *Proceedings of the SPIE (44th Annual Meeting)*, **3813**: 188-195.
- [427] Simonoff J 1996 *Smoothing Methods in Statistics*, Springer, New York, New York, USA.

- [428] Sinha D, Sinha P, Dougherty ER & Batman S 1997 'Design and analysis of fuzzy morphological algorithms for image processing', *IEEE Transactions on Fuzzy Systems*, **5**(4):570-584.
- [429] Siu KKW, Butler SM, Beveridge T, Gillam JE, Hall CJ, Kaye AH, Lewis RA, Mannan K, McLoughlin G, Pearson S, Round AR, Schültke E, Webb GI & Wilkinson SJ 2005 'Identifying markers of pathology in SAXS data of malignant tissues of the brain', *Nuclear Instruments and Methods in Physics Research A* **548**(1-2): 140-146.
- [430] Small CG 1996 *The Statistics Theory of Shape*, Springer-Verlag, New York, New York, USA.
- [431] Sofou A & Maragos P 2003 'PDE-based modeling of image segmentation using volumic flooding', *ICIP 2003. Proceedings 2003 International Conference on Image Processing*, **2**:II-431-434.
- [432] Speller RD & Horrocks JA 1991 'Photon scattering -a 'new' source of information in medicine and biology', *Physics in Medicine and Biology*, **36**: 1-6.
- [433] Sreenivasan K 1991 'Fractals and multifractals in fluid turbulence', *Annual Reviews Fluid Mechanics*, **23**: 539-600.
- [434] Starck J-L & Bijaoui A 1994 'Filtering and deconvolution by the wavelet transform', *Signal Processing*, **35**(3): 195-211.
- [435] Starck J-L, Candès EJ & Donoho DL 2002 'The curvelet transform for image denoising', *IEEE Transactions on Image Processing*, **11**(6):670-684.
- [436] Stone HS, Orchard MT, Chang E-C & Martucci SA 2001 'A fast direct Fourier-based algorithm for subpixel registration of images', *IEEE Transactions on Geoscience & Remote Sensing*, **39**(10):2235-2243.
- [437] Stošić T & Stošić BD 2006 'Multifractal analysis of human retinal vessels', *IEEE Transactions on Medical Imaging*, **25**(8):1101-1107.
- [438] Strang G 1999 'The discrete cosine transform', *SIAM Review*, **41**(1):135-147.
- [439] Strauss O & Comby F 2005 'Fuzzy morphology for omnidirectional images', *IEEE Conference on Image Processing, 2005. ICIP 2005*, **2**:141-144.

- [440] Sturges HA 1926 'The choice of a class interval', *The Journal of the American Statistical Association*, **21**(153):65-66.
- [441] Suhonen H, Fernández M, Serimaa R & Suortti P 2005 'Simulation of small-angle x-ray scattering from collagen fibrils and comparison with experimental patterns', *Physics in Medicine and Biology*, **50**: 5401-5416.
- [442] Sullivan F & Hunt F 1988 'How to estimate capacity dimension', *Nuclear Physics B (Proc. Suppl.)*, **5A**: 125-128.
- [443] Suortti P & Thomlinson W 2003 'Medical applications of synchrotron radiation', *Physics in Medicine and Biology*, **48**(13): R1-R35.
- [444] Suri JS 'White matter/grey matter boundary segmentation using geometric snakes: A fuzzy deformable model' in Murshed SSN & Kropatsch W (eds) 2001, *ICAPR 2001, LNCS 2013*, Springer-Verlag, Berlin Heidelberg, Germany, pp. 331-338
- [445] Suri JS, Wilson DL, Laxminarayan S 2005 *Handbook of Biomedical Image Analysis*, Kluwer Academic/Plenum Publishers, New York, New York, USA.
- [446] Tang L, Wang H & Qi B 2007 'A new fuzzy logic image de-noising algorithm based on gradient detection', *Fourth International Conference on Fuzzy Systems and Knowledge Discovery*, **2**:103-107.
- [447] Tao C-W, Thompson WE & Tour JS 1993 'A fuzzy if-then approach to edge detection', *Second IEEE International Conference on Fuzzy Conference on Fuzzy Systems*, **2**:1356-1360.
- [448] Tartari A, Casnati E, Bonifazzi C & Baraldi C 1997 'Molecular differential cross sections for x-ray coherent scattering in fat and polymethyl methacrylate', *Physics in Medicine and Biology*, **42**: 2551-2560.
- [449] Tartari A, Taibi A, Bonifazzi C & Baraldi C 2002 'Updating of form factor tabulations for coherent scattering of photons in tissues', *Physics in Medicine and Biology*, **47**: 163-175.
- [450] Tarter ME & Kromal RA 1970 'On multivariate density estimates based on orthogonal expansions', *The Annals of Mathematical Statistics*, **41**(2):718-722.

- [451] Tarter ME & Kronmal RA 1976 'An introduction to the implementation and theory of nonparametric density estimation', *The American Statistician*, **30**:105-112.
- [452] Teboul S, Blanc-Feraud L, Aubert G & Barlaud M 1998 'Variational approach for edge-preserving regularisation using coupled PDE's', *IEEE Transactions on Image Processing*, **7**(3):387-397.
- [453] Tek H & Kimia BB 'Curve evolution and mathematical morphology' in Heijmans HJAM & Roerdink JBTM (eds) 1998 *Mathematical Morphology and its Applications to Image and Signal Processing*, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- [454] Terrell GR & Scott DW 1992 'Variable kernel density estimation', *The Annals of Statistics*, **20**(3):1236-1265.
- [455] Tessier Y 1993 'Multifractal objective analysis, rain and clouds', *Doctoral Thesis*, McGill University, Montreal, Quebec, Canada.
- [456] Tessier Y, Lovejoy S & Schertzer D 1993 'Universal multifractals: theory and observations for rain and clouds', *Journal of Applied Meteorology*, **32**(2):223-250.
- [457] Tessier Y, Lovejoy S, Hubert P, Schertzer D & Pecknold S 1996 'Multifractal analysis and modeling of rainfall and river flows and scaling, casual transfer functions', *Journal of Geophysical Research*, **101**(D/21):427-440.
- [458] Theiler J 1990 'Estimating fractal dimension', *Journal of the Optical Society of America A*, **7**(6):1055-1073.
- [459] Theodorakou C & Farquharson MJ 2008 'Human soft tissue analysis using x-ray or gamma-ray techniques', *Physics in Medicine and Biology*, **53**: R111-R149.
- [460] Tizhoosh HR 2002 'Fast fuzzy edge detection', *Proceedings of the 2002 Annual Meeting of the North Americas Fuzzy Information Processing Society*, pp.239-242.
- [461] Torrence C & Campo GP 1998 'A practical guide to wavelet analysis', *Bulletin of the American Meteorological Society*, **79**(1):61-78.
- [462] Trivedi MM & Bezdek JC 1986 'Low-level segmentation of aerial images with fuzzy clustering', *IEEE Transactions on Systems, Man & Cybernetics*, **16**(4):589-598.

- [463] Tsai A, Yezzi A & Willsky AS 2001 'Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation and magnification', *IEEE Transactions on Image Processing*, **10**(8):1169-1186.
- [464] Tschumperle D & Deriche R 2002 'Diffusion PDEs on vector-valued images', *IEEE Signal Processing Magazine*, **95**(5):16-25.
- [465] Udupa JK & Samarsekera S 1996 'Fuzzy connectedness and object definition: theory, algorithms and applications in image segmentation', *Graphical Models and Image Processing*, **58**(3):246-261.
- [466] Udupa JK, Wei L, Samarasekera S, Miki Y, Van Buchem MA & Grossman RI 1997 'Multiple sclerosis lesion quantification using fuzzy-connectedness principle', *IEEE Transactions on Medical Imaging*, **16**(5):598-609.
- [467] Vapnik VN 1995 *The Nature of Statistical Learning Theory*, Springer, Berlin, Germany.
- [468] Varshney LR 2004 *Despeckling synthetic aperture radar imagery using the contourlet transform*, Technical Report, School of Electrical and Computer Engineering, Cornell University, Ithaca, New York, USA.
- [469] Vicsek T 1989 *Fractal Growth Phenomena*, World Scientific, Singapore.
- [470] Vidakovic B 1999 *Statistical Modeling by Wavelets*, John Wiley & Sons, New York, New York, USA.
- [471] Von Mises R 1954 La distribution de la plus grande de n valeurs, *Selected Papers, Volume II*, 271-294, American Mathematical Society, Providence, Rhode Island, USA .
- [472] Voss RF 1991 'Random fractals: characterisation and measurement' in Family F & Vicsek T (eds) 1991 *Dynamics of Fractal Surfaces*, World Scientific Publishing, Singapore.
- [473] Wahba G 1981 'Data-based optimal smoothing of orthogonal series density estimates', *The Annals of Statistics*, **9**(1):146-156.
- [474] Walter GG 1994 *Wavelets and Other Orthogonal Systems with Applications*, CRC Press, Boca Raton, Florida, USA.

- [475] Wand MP, Marron JS & Ruppert D 1991 'Transformations in Density Estimation', *Journal of the American Statistical Association*, **86**(414):343-353.
- [476] Wand MP & Jones MC 1995 *Kernel Smoothing*, Chapman & Hall/CRC, London, UK.
- [477] Wand MP 1997 'Data-based choice of histogram bin width', *American Statistician*, **51**(1):59-64.
- [478] Wang W, Wyckoff JB, Frohlich VC, Oleynikov Y, Hüttelmaier S, Zavadil J, Cermak L, Bottinger EP, Singer RH, White JG, Segall JE, Condeelis JS 2002 'Single cell behaviour in metastatic primary mammary tumours correlated with gene expression patterns revealed by molecular profiling' *Cancer Research*, **62**: 6278-6288.
- [479] Wang Y, Gu X, Chan TF, Thompson PM, Yau S-T 2004 'Intrinsic brain surface conformal mapping using a variational method', *Proceedings of the SPIE*, **5370**(1):241-252.
- [480] Watson GS 1969 'Density estimation by orthogonal series', *The Annals of Mathematical Statistics*, **40**(4):1496-1498.
- [481] Weeratunga SK & Kamath C 'A comparison of PDE-based non-linear anisotropic diffusion techniques for image denoising' in Dougerty ER, Astola JT & Egiazarian KO (eds) 2003 *Image Processing: Algorithms and Systems II*, SPIE-The International Society for Optical Engineering, Santa Clara, California, USA.
- [482] Weickert J 2001 'Efficient image segmentation using partial differential equations and morphology', *Pattern Recognition*, **34**(9):1813-1824.
- [483] Weisstein EW 2003 *CRC Concise Encyclopedia of Mathematics*, (2nd edn), Chapman & Hall/CRC, Boca Raton, Florida, USA.
- [484] Welk M, Theis D, Brox T & Weickert J 'PDE-based deconvolution with forward-backward diffusivities and diffusion tensors' in Kimmel R, Sochen N & Weickert J (eds) 2005 *Scale Space and PDE Methods in Computer Vision*, Springer-Verlag, Berlin Heidelberg, Germany.
- [485] Welland GV (ed) 2003 *Beyond Wavelets*, Academic Press, San Diego, California, USA.
- [486] Wess TJ 2005 'Collagen fibril form and function', *Advances in Protein Chemistry*, **70**:341-374.

- [487] Westmore MS, Fenster A & Cunningham 1996 'Angular-dependent coherent scatter measured with a diagnostic x-ray image intensifier-based imaging system', *Medical Physics*, **23**(5): 723-733.
- [488] Whittaker ET 1915 'On functions which are represented by the expansion of interpolating theory', *Proceedings of the Royal Society of Edinburgh*, **35**: 181-194.
- [489] Whittaker JM 1929 'The Fourier theory of Cardinal functions', *Proceedings of the Mathematical Society of Edinburgh*, **1**:169-176.
- [490] Whittaker JM 1935 *Interpolating Function Theory*, Cambridge University Press, Cambridge, United Kingdom.
- [491] Whittle P 1958 'On the smoothing of probability density functions', *Journal of the Royal Statistical Society Series B*, **20**(2):334-343.
- [492] Wilkinson SJ, Rogers KD & Hall CJ 2006 'Model fitting in two dimensions to small angle diffraction patterns from soft tissue', *Physics in Medicine and Biology*, **51**: 1819-1830.
- [493] Wilkinson SJ, Rogers KD, Hall CJ, & Round AR 2007 'Small angle scatter imaging from wide beam diffraction patterns', *Physics in Medicine & Biology*, **52**(9):2409-2423.
- [494] Winkler G 1995 *Image Analysis, Random Fields, and Dynamic Monte Carlo Methods: A Mathematical Introduction*, Springer-Verlag, Berlin, Germany.
- [495] Wolfram Research, Inc., 2008 *Mathematica*, Version 7.0, Champaign, Illinois, USA.
- [496] Wu J, Yin ZP & Xiang Y 2007 'The fast multilevel fuzzy edge detection of blurry images', *IEEE Signal Processing Letters*, **14**(5):344-347.
- [497] Yaffe MJ, Byng JW & Boyd NF 'Quantitative image analysis for estimation of breast cancer risk' in Bankmans IN (ed) 2000 *Handbook of Medical Imaging and Analysis*, (2nd edn), Academic Press, San Diego, California, USA.
- [498] Yezzi Jr A, Kichenassamy S, Kumar A, Olver P & Tannenbaum A 1997 'A geometric snake model for segmentation of medical imagery', *IEEE Transactions on Medical Imaging*, **16**(2):199-209.

- [499] Yezzi Jr AJ & Prince JL 2003 'An Eulerian PDE approach for computing tissue thickness', *IEEE Transactions on Medical Imaging*, **22**(10):1132-1339.
- [500] Yezzi A, Zöllei L & Kapur T 2003 'A variational framework for integrating segmentation and registration through active contours', *Medical Image Analysis*, **7**(2):171-185.
- [501] You YL & Kaveh M 2000 'Fourth-order partial differential equations for noise removal', *IEEE Transactions on Image Processing*, **9**(10):1723-1730.
- [502] Young HD & Freedman RA 1996 *University Physics* (9th edn), Addison-Wesley, New York, New York, USA.
- [503] Zadeh L 1965 'Fuzzy sets', *Information Control*, **8**:338-353.
- [504] Zahn CT & Roskies RZ 1972 'Fourier descriptors for plane-closed contours', *IEEE Transactions on Computers*, **21**(3):269-281.
- [505] Zhang L & Bao P 2002 'Edge detection by scale multiplication in the wavelet domain', *Pattern Recognition Letters*, **23**(14): 1771-1784.
- [506] Zhang X & Zhang C 2007 'Satellite cloud image denoising and enhancement by fuzzy wavelet neural network and genetic algorithm in the curvelet domain', *Lectures Notes in Computer Science*, **4688**:389-395.
- [507] Zhu M & Martinez AM 2006 'Subclass discriminant analysis', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **28**(8):1274-1286.