



## City Research Online

### City, University of London Institutional Repository

---

**Citation:** Reyes-Aldasoro, C. C. & Bhalerao, A. (2006). The Bhattacharyya space for feature selection and its application to texture segmentation. *Pattern Recognition*, 39(5), pp. 812-826. doi: 10.1016/j.patcog.2005.12.003

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/4313/>

**Link to published version:** <https://doi.org/10.1016/j.patcog.2005.12.003>

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

---

---



# The Bhattacharyya Space for Feature Selection and its application to Texture Segmentation

C. C. Reyes-Aldasoro <sup>a,\*</sup> A. Bhalerao <sup>b</sup>

<sup>a</sup>*Tumour Microcirculation Group,  
Academic Unit of Surgical Oncology,  
The University of Sheffield, Sheffield, S10 2JF, UK*

<sup>b</sup>*Department of Computer Science,  
University of Warwick, Coventry, CV4 7AL, UK*

---

## Abstract

A feature selection methodology based on a novel Bhattacharyya Space is presented and illustrated with a texture segmentation problem. The Bhattacharyya Space is constructed from the Bhattacharyya distances of different measurements extracted with sub-band filters from training samples. The marginal distributions of the Bhattacharyya Space present a sequence of the most discriminant sub-bands that can be used as a path for a wrapper algorithm. When this feature selection is used with a multiresolution classification algorithm on a standard set of texture mosaics, it produces the lowest misclassification errors reported.

*Key words:* Feature Selection, Bhattacharyya distance/space, Texture Segmentation

---

## 1 Introduction

The problems of feature selection and texture segmentation have been studied by pattern recognition, image processing and computer vision researchers for a number of years and they continue to be of interest due to the wealth of applications and also the desire to produce accurate results at a low computational cost.

---

\* Corresponding author.

*Email address:* `c.reyes@sheffield.ac.uk` (C. C. Reyes-Aldasoro).

Feature selection is a fundamental pre-processing step in any classical pattern recognition problem, and the growth of computer storage and power has enabled more complex measurements on larger input data which result in correspondingly large numbers of high dimensional features [21,15]. Therefore, methods that can select appropriate and compact subsets of features are vital to the accuracy and efficiency of any subsequent classification step.

The feature selection and extraction problem considers the mathematical tools for reducing the dimensionality of a Measurement Space [16], which is sometimes called Pattern Representation [23] or Feature Space. The problem faced is that of selecting a feature subset which will reduce the complexity at the classifier without affecting its performance. The reduced subset can be obtained in two different ways: *feature selection* or *feature extraction*. In feature selection, a set of the original measurements is discarded and the ones that are selected, which will be the most useful ones, will constitute the *Feature Space*. In contrast, the combination of a series of measurements in a linear or non-linear mapping to a new reduced dimensionality is called feature extraction. *Feature construction* [29,26] relies on additional information, which will not be assumed in the present work, to add new features in order to simplify hypothesis search.

Ideally, the best way to obtain a reduced feature set is to test every combination of measurements through the classifier. For  $N_i$  measurements, there will be  $O(2^{N_i})$  different solutions, which yield computations impractical even for small number of measurements. Branching techniques [23] can obtain optimal solutions but they are still computationally intensive. It is necessary then to settle for sub-optimal solutions that will not analyse the whole space of combinations exhaustively. The simplest of these solutions can be placed into two groups called *forward selection*, and *backward elimination* (which are both particular cases of the *plus l - take away r algorithm*). In forward selection, a search begins with an empty set of features, and elements are sequentially included at a classifier, the selection will depend of an individual best measurement. In backward elimination the starting state is the full set of features, and measurements are discarded one by one. The process of selection or elimination continues up to a certain state where an evaluation criterion is satisfied and a final subset is reached. The selection implies that if each of the elements of the subset is forwarded sequentially to a classifier, then we expect to improve the classification, but if we were to continue with any other element not in the subset, then there would be a degradation of the results.

Feature extraction will use all the dimensions of the measurement space and map it to a lower dimensional space, where the new features will contain the useful information through a projection that will ignore redundant and irrelevant information. Perhaps the most common feature extraction method is *Principal Components Analysis* (PCA) where the new features are uncorre-

lated and these are the projections onto axes that maximise the variances of the data. As well as making each feature linearly independent, PCA allows the ranking of features according to the size of the global covariance in each principal axis from which a ‘subspace’ of features can be presented to a classifier. However, while this eigenspace method is effective in many cases, it requires the computation of all the features for given data. In some of the applications presented in this work, the measurement space need only be generated for a set of training samples. These will be used to determine a feature space and then only the required features are obtained for the whole data set considerably reducing the computational effort.

Image texture, as well as feature selection, has been well studied in the past decades and because of its application in many areas such as of Crystallography [46], Stratigraphy, [6,36], Medical Imaging, (Magnetic Resonance Imaging (MRI) [27,41,25], Ultrasound [53] or Computed Tomography (CT) [19,43]), or content-based image retrieval [28] continues to be of interest and many papers on texture extraction, segmentation and classification are still published every year. In some cases, texture has been analysed not only in 2D but also in 3D [2,38,39].

Many different approaches for 2D texture measurement generation, classification and segmentation have been reported, for example: [33,37,47,50]. One of the most common approaches of 2D image texture description is the use of Haralick’s co-occurrence analysis, first published in the 1970s [18,17] and still widely used today. Cross and Jain [9] and Chellapa and Jain [7] reported with some success on statistical approaches using Markov Random Fields for the modelling of texture. Jain and Farrokhnia [20] observed the spectral energy of textures with Gabor filters. Since texture can be scale dependent, wavelets and other multiresolution techniques have been widely used by Unser [48] and others [3,51,49,42]. In a recent and thorough study, Randen and Husøy [35] have compared different filter-based approaches against a set of natural textures from the classical Brodatz Album [4] and other databases [44,32]. The composite images contain different natural textures that were captured under different illumination conditions and with different equipment, but were selected to be visually stationary. Each texture has been globally histogram equalised and they have the same mean value so that they spread the same range of grey levels. Some of the masks that were used to form these images contain triangular and circular shapes which are harder to segment than squares or rectangles. Randen’s images, which are fairly hard to classify even by eye, are becoming a benchmark for assessing different segmentation algorithms, [30,31,33].

The nine texture images segmented in this work correspond to figure 11 in [35] and are presented in figure 1. Figures (a) to (e) consist of 5 different textures in images with size  $256 \times 256$  pixels, (f) and (g) have 16 textures and are  $512 \times 512$

pixels, (h) and (i) have 10 textures and  $640 \times 256$  pixels. In their study, Randen and Husøy show that filtering methods outperform co-occurrence texture measures but vary in their computational cost, number, type and decomposition of features used and ease of implementation with best overall results being obtained by multiresolution wavelet and quadrature-mirror filters.

The rest of the paper is organised as follows. In section 2 the measurement space is generated by sub-band filtering with and an Orientation Pyramid (OP). Two classification strategies are then presented. First, a single resolution to compare the quality of the measurement space with those presented by Randen, and then a multiresolution algorithm that can easily outperform the single resolution. Section 3 introduces the Bhattacharyya distances and the presents the novel Bhattacharyya space as the basis of a feature selection algorithm and further improvements are demonstrated. Section 4 presents comparative results on 9 multitextured images. Finally conclusions are presented.

## 2 Methodology

### 2.1 Feature Extraction: Sub-band filtering using an Orientation Pyramid (OP)

Certain characteristics of signals in the spatial domain such as periodicity are quite distinctive in the frequency or Fourier domain. If the data contain textures that vary in orientation and frequency, then certain filter sub-bands will contain more energy than others.

Wilson and Spann [52] proposed a set of operations that subdivide the frequency domain of an image into smaller regions by the use of two operators *quadrant* and *centre-surround*. By combining these operators, it is possible to construct different tessellations of the space, one of which is the Orientation Pyramid (OP) (Figure 2). A band-limited filter based on truncated Gaussians is used to approximate the finite prolate spheroidal sequences used in [52]. The filters are real functions which cover the Fourier half-plane. Since the Fourier transform of a real signal is symmetric, it is only necessary to use a half-plane or a half-volume to measure sub-band energies. A description of the sub-band filtering with the OP method follows.

Throughout this work, we will consider an image,  $\mathcal{I}$ , represented as a function that assigns a grey tone to each pair of co-ordinates [18]:

$$L_r \times L_c; \mathcal{I} : L_r \times L_c \rightarrow G, \quad (1)$$

where  $N_r \times N_c$  are the dimensions of rows and columns,  $L_r = \{1, 2, \dots, r, \dots, N_r\}$ ,

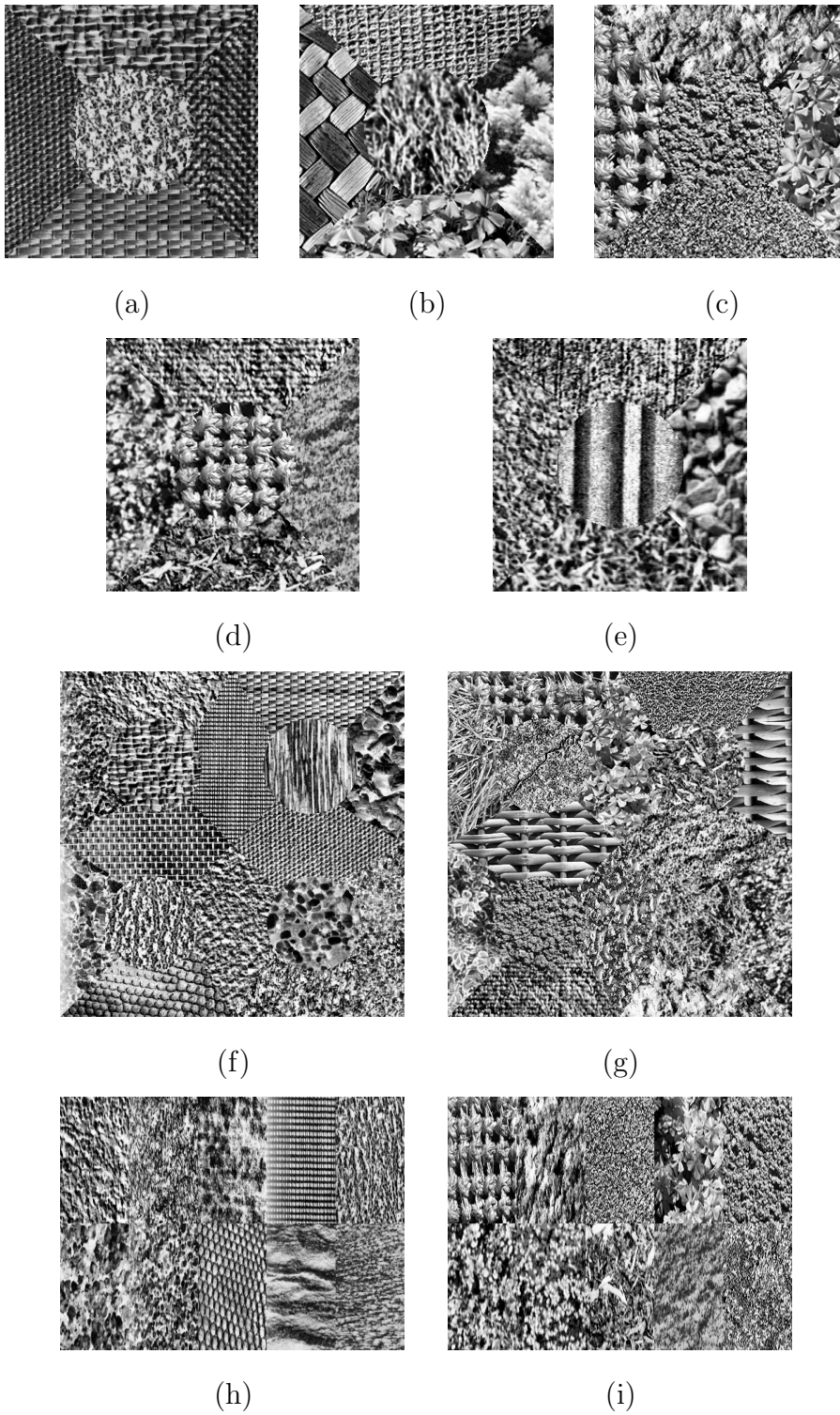


Fig. 1. Composite texture images arranged by Randen and Husøy [35].

$L_c = \{1, 2, \dots, c, \dots, N_c\}$  are the spatial domains of each dimension,  $L_r \times L_c$  is the domain of image, and  $G = \{1, 2, \dots, g, \dots, N_g\}$  is the set of  $N_g$  grey levels; the co-domain of the image.

The centred Fourier transform of  $\mathcal{I}$ ,  $\mathcal{I}_\omega = \mathcal{F}[\mathcal{I}]$ , can be subdivided into a set of  $i$  non-overlapping regions  $L_r^i \times L_c^i$  of dimensions  $N_r^i, N_c^i$ . The OP tessellation involves a set of 7 filters, one for the low pass region and six for the high pass (Figure 2 (a)). The  $i$ -th filter  $F_\omega^i$  in the Fourier domain ( $F_\omega^i = \mathcal{F}[F^i]$ ) is related to the  $i$ -th subdivision of the frequency domain as:

$$L_r \times L_c; F_\omega^i : \begin{cases} L_r^i \times L_c^i & \rightarrow \mathcal{G}_a(\mu^i, \Sigma^i), \\ (L_r^i \times L_c^i)^c & \rightarrow 0 \end{cases} \quad \forall i \in OP,$$

where  $\mathcal{G}_a$  describes a Gaussian function, with parameters  $\mu^i$ , the centre of the region  $i$ , and  $\Sigma^i$  is the co-variance matrix that will provide a cut-off of 0.5 at the limit of the band. The measurement space  $S$  in its frequency and spatial

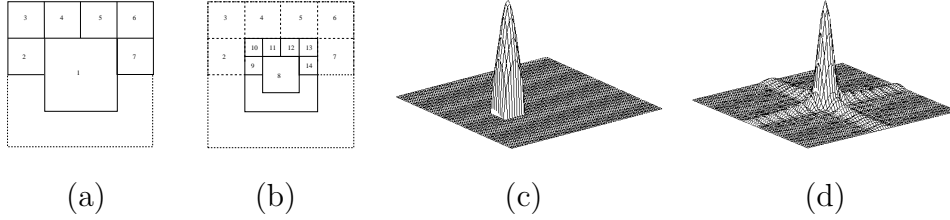


Fig. 2. Orientation Pyramid (OP) tessellation: (a) order 1, (b) order 2. Band-limited 2D Gaussian filter: (c) Frequency domain  $|F_\omega^i|$ , (d) Magnitude of spatial domain  $|F^i|$ .

domains is then defined as:

$$\begin{aligned} S_w^i(\rho, \kappa) &= F_\omega^i(\rho, \kappa) \mathcal{I}_\omega(\rho, \kappa) \\ S^i &= |\mathcal{F}^{-1}[S_w^i]|, \end{aligned} \quad (2)$$

where  $(\rho, \kappa)$  are the co-ordinates in the Fourier domain. The OP can be further subdivided, at the next level the coordinates  $(L_r^1(1) \times L_c^1(1))$  will become  $(L_r(2) \times L_c(2))$  with dimensions  $N_r(2) = \frac{N_r(1)}{2}, N_c(2) = \frac{N_c(1)}{2}$ . (Figure 2 (b)). More levels can be obtained provided that the image has the required dimensions. It is assumed that  $N_r(1) = 2^a, N_c(1) = 2^b$  so that the results of the divisions are always integer values. To illustrate the OP on a textured image, one of Randen's images is filtered and presented in Figure 3.

## 2.2 Classification of the Measurement Space

Partitioning of the measurement space can be considered as a mapping operator  $\lambda : S \rightarrow \{1, 2, \dots, N_k\}$ , where the clusters or classes are  $\lambda^{-1}(1), \lambda^{-1}(2)$ , etc., and these are unknown. Then, for every element  $x \in S$ ,  $\lambda_a$  will be an estimator for  $\lambda$  where, for every class, there is a point  $\{a_1, a_2, \dots\} \in S$  such that these points define hyperplanes perpendicular to the chords connecting them,



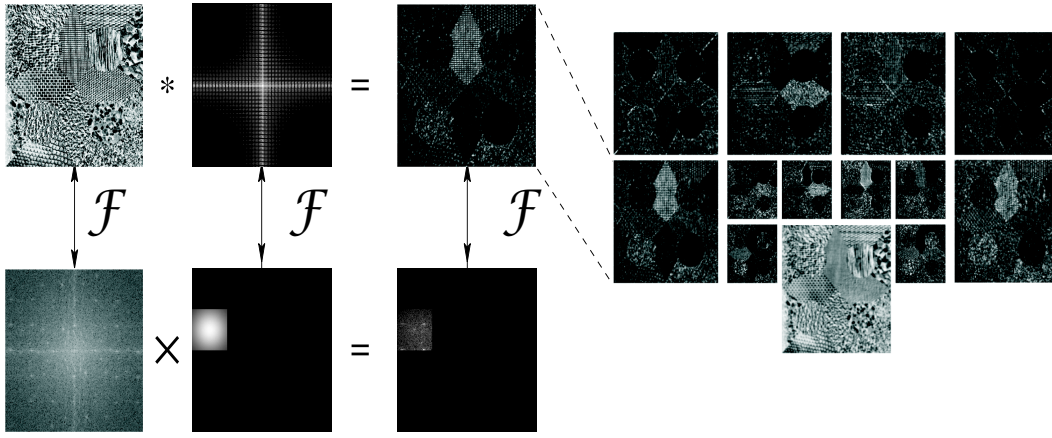


Fig. 3. A graphical example of sub-band filtering. The top row corresponds to the spatial domain and the bottom row the Fourier domain. A 16-texture (figure 1 (f)) is filtered with a sub-band filter with a particular frequency and orientation by a product in the Fourier domain, which is equivalent to a convolution in the spatial domain. The filtered image becomes one measurement of the space,  $S^2$  in this example.

and split the space into regions  $\{R_1, R_2, \dots\}$ . These regions define the mapping function  $\lambda_a : S \rightarrow \{1, 2, \dots, N_k\}$  by  $\lambda_a(x) = K$  if  $x \in R_K, K = 1, 2, \dots, N_K$ . This partitioning should minimise the Euclidean distance from the elements of the space to the points  $a$ , expressed by [11]:

$$\rho(a_1, a_2, \dots) = \sum_{x \in (L_r \times L_c)} \min_{1 \leq j \leq N_k} \|S(x) - a_j\|. \quad (3)$$

The measure of closeness of the estimator  $\lambda_a$  to  $\lambda$  defines a misclassification error by  $\epsilon[\lambda_a] = P(\lambda_a(x) \neq \lambda(x))$ , for an arbitrary point  $x \in S$  in the space.

If the values of the points  $a_k$  are known, or there is a way of estimating these from training data, the classification procedure is *supervised*, otherwise it is *unsupervised*. For this work, the points in the measurement space  $a_k$  were obtained by filtering separate training data with the OP. Once the measurement space  $S$  is calculated for every training image, the average can be used as an estimate of the mean of the class:  $\hat{a}_k$ .

Table 1 compares the results of the sub-band filtering with 35 measurements (order 5 of the OP) and a  $13 \times 13$  Gaussian local energy function (LEF) (for more details of the effect of the LEF, see [38]) with different measurement extraction techniques. These results confirm that sub-band filtering with an OP can extract textural measurements that are as good as those presented by Randen.

Table 1

Comparative misclassification results (%) of the natural textures (Table 3 in [35]) and OP sub-band filtering. Best results are in **bold**.

Misclassification (%)	Figures									
Measurement	a	b	c	d	e	f	g	h	i	Average
Laws	9.7	25.7	32.4	27.3	25.7	48.3	54.3	41.9	37.8	33.68
Ring/Wedge	14.6	35.5	28.9	35.5	22.4	43.8	67.8	44.5	48.3	37.92
Dyadic Gabor	10.7	34.8	22.6	25.2	24.6	60.1	58.2	32.3	47.9	35.16
Gabor Banks	<b>8.2</b>	34.0	25.8	36.9	28.4	54.8	71.5	39.7	54.8	39.34
DCT	13.2	27.0	25.5	37.8	22.6	40.9	49.0	38.2	33.0	31.91
Daubechies 4	8.7	22.8	25.0	23.4	21.8	38.2	45.2	40.9	30.1	28.46
fl6b	8.7	<b>18.9</b>	23.3	<b>18.4</b>	<b>17.2</b>	36.4	<b>41.7</b>	39.8	28.5	<b>25.88</b>
Co-occurrence	9.9	27.0	26.1	51.1	35.7	49.6	55.4	35.3	49.1	37.69
AR	19.6	19.4	23.0	23.9	34.0	58.0	46.4	56.7	<b>28.7</b>	34.41
Average	11.5	27.2	25.9	31.1	24.7	47.8	54.4	41.0	39.8	33.71
OP	9.0	31.7	<b>20.6</b>	20.7	<b>17.2</b>	<b>32.7</b>	49.5	<b>27.9</b>	39.5	27.6

### 2.3 Multiresolution Classification

A multiresolution classification strategy can exploit the inherent multiscale nature of texture and better results can be achieved. The multiresolution procedure consists of three main stages: *climb*, *decide* and *descend*.

The climbing stage represents the decrease in resolution of the data by means of averaging a set of neighbours on one level (*children* elements or nodes) up to a *parent* element on the upper level. Two common climbing methods are the Gaussian Pyramid [5] and the Quad tree *QT* ([14,40,45]). In our implementation we used the *QT* structure. The decrease in resolution correspondingly reduces the uncertainty in the elements' values since they tend toward their mean. In contrast, the positional uncertainty increases at each level [52].

At the highest level, the new reduced space can be classified either in a supervised or unsupervised scheme as it was described before.

To regain full spatial resolution at the lowest level of the tree, the classification at the highest level has to be propagated downward. The propagation implies that every parent bequeaths: (a) its class value to 4 children and; (b) the attribute of being or not being in a boundary. As the classification is propagated, a spatial restoration process can be performed at every level to reduce the uncertainty in the spatial position. This typically implies an interaction of an element with its neighbours to eliminate isolated pixels and a selective smoothing can be performed with butterfly filters. Butterfly filters (*BF*) [42] are orientation-adaptive filters, that consist of two separate sets or *wings* with a pivot element between them. It is the pivot element  $x = (r, c)$  which is modified as a result of the filtering. Each of the wings will have a roughly triangular shape, which resembles a butterfly and they can be regarded as two separate sets of *anisotropic cliques*, arranged in a steerable orientation. The elements covered by each of the wings are included in the filtering process while the values of the elements along the boundary (which are presumed to

have greater uncertainty) and the pivot,  $x$ , are not included in the smoothing process. The use of  $BF$  outperforms other multiresolution schemes such as Markov Random Fields and they can be extended to 3D [38].

### 3 Feature Selection using the Bhattacharyya space

In the previous section, a measurement space was generated by sub-band filtering the textured images. This space will consist of a number of dimensions, which could equally be generated by Gabor filters, features of the co-occurrence matrix or wavelets, and not all the dimensions will contribute to the discrimination of the different textures that compose the original data. Besides the discrimination power that some features have, there is also a complexity issue related to the number of features selected. Another advantage of selecting a subset of features is that they can provide a better understanding of the underlying process that generated the data [15].

One of the most common methods [10] of forward selection is the *wrapper approach* [24]. This approach uses the error rate of a classifier itself as the criterion to evaluate the features selected, it proposes greedy selection, either as *hill climbing*, or *best first* as search algorithms and treats the measurements as a search space organisation, a representation where each state represents a measurement subset. For  $N_i$  measurements, there are  $N_i$  bits in each state indicating the presence (1) or absence (0) of the measurement. The state  $\{0, 0, \dots, 0\}$ , the empty set will be the initial state for forward selection, and  $\{1, 1, \dots, 1\}$  will describe the whole measurement space (initial state for backward elimination). Figure 4 shows a 4-measurement state space where forward and backward selection processes have been identified. Each of the links will represent a single measurement added (continuous line) or deleted (dashed line).

The process of wrapper selection with a hill climbing search follows the sequence:

- (1) Start with an empty set of features  $v \leftarrow \{0, 0, \dots, 0\}$ .
- (2) Expand  $v$ : generate new states by adding a single feature from  $v$ . In the example of Figure 4 (a) the children of  $v$  are  $\{1, 0, 0, 0\}$ ,  $\{0, 1, 0, 0\}$ ,  $\{0, 0, 1, 0\}$ ,  $\{0, 0, 0, 1\}$ .
- (3) Apply the evaluation function  $\lambda$  (that is, the classifier) to each child  $w$  of  $v$ .
- (4) Let  $v' =$  the child with the highest evaluation  $\lambda(w)$ .
- (5) If  $\lambda(v') > \lambda(v)$  then  $v \leftarrow v'$  and go to 2, else finish with  $v$  as a final subset.

The previous algorithm is the basic presentation and it can easily be varied; for example, different ways of expanding  $v$  rather than just considering every child can be used. It is important to bear in mind two issues: one is that hill climbing can lead to local optima, and the other is that the strength of the algorithm, the use of the classifier in the selection process instead of other evaluation functions, is at the same time its weakness, since the classification process can be slow.

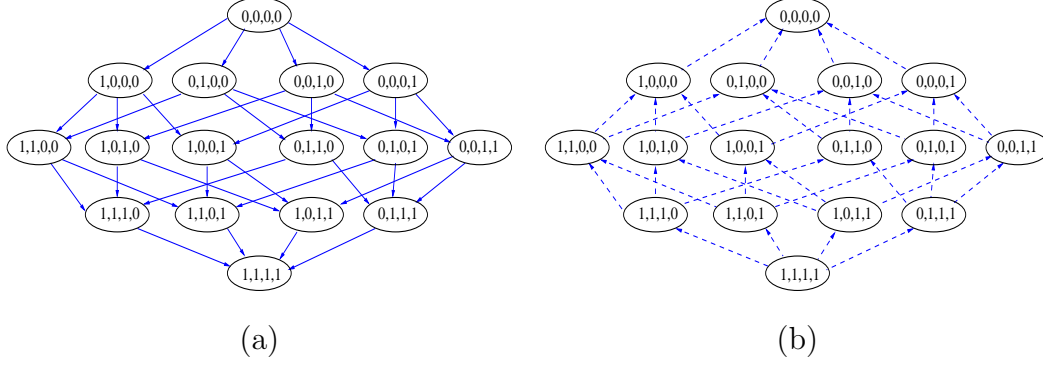


Fig. 4. State Space for sequential selection. Each node is connected to nodes that have one measurement added or deleted (a) Forward selection (b) Backward selection.

One way to avoid the evaluation of each child of the current state will be proposed below. The *Bhattacharyya Space* is presented as a method that provides a ranking for the measurements based on the discrimination of a training set. This ranking process provides a single route to evaluate and therefore, the number of classifications, which will still be done for every feature added to the classifier, is significantly reduced. Since this method a pre-processing step, and is calculated over training data (of small size compared to the whole data set), a heuristic solution to avoid being trapped by local optima is also proposed.

### 3.1 The Bhattacharyya distance

In order to obtain a quantitative measure of *how separable* are two classes, a distance measure is required. With the assumption of underlying distributions a *probabilistic distance* a distance can be easily extracted from some parameters of the data. Kailath [22] compared the Bhattacharyya Distance and the Divergence (Kullback-Leibler), and observed that Bhattacharyya yields better results in some cases while in other cases they are equivalent. In a recent study [1], a number of measures; Bhattacharyya, Euclidean, Kullback-Leibler, Fisher, have been studied for image discrimination and it was concluded that the *Bhattacharyya distance* [13] is the most effective texture discrimination for sub-band filtering schemes.

In its simplest formulation, the Bhattacharyya distance between two classes can be calculated from the variance and mean of each class in the following way [8]:

$$D_B(k_1, k_2) = \frac{1}{4} \ln \left\{ \frac{1}{4} \left( \frac{\sigma_{k_1}^2}{\sigma_{k_2}^2} + \frac{\sigma_{k_2}^2}{\sigma_{k_1}^2} + 2 \right) \right\} + \frac{1}{4} \left\{ \frac{(\mu_{k_1} - \mu_{k_2})^2}{\sigma_{k_1}^2 + \sigma_{k_2}^2} \right\} \quad (4)$$

where:  $D_B(k_1, k_2)$  is the Bhattacharyya distance between  $k_1 - th$  and  $k_2 - th$  classes,  $\sigma_{k_1}$  is the variance of the  $k_1 - th$  class,  $\mu_{k_1}$  is the mean of the  $k_1 - th$  class, and  $k_1, k_2$  are two different training classes.

For the multidimensional distance, the variances are replaced by co-variance matrices and the means become vectors [13]:

$$D_B(k_1, k_2) = \frac{1}{2} \ln \left[ \frac{|\frac{1}{2}(\Sigma_{k_1} + \Sigma_{k_2})|}{\sqrt{|\Sigma_{k_2}| |\Sigma_{k_1}|}} \right] + \frac{1}{4} (\mu_{k_1} - \mu_{k_2})^T [\Sigma_{k_1} + \Sigma_{k_2}]^{-1} (\mu_{k_1} - \mu_{k_2}) \quad (5)$$

The Mahalanobis distance used in Fisher LDA is a particular case of the Bhattacharyya, when the variances of the two classes are equal, this would eliminate the first term of the distance. This term depends solely of the variances of the distribution. If the variances are equal this term will be zero, and it will grow as the variances are different. The second term, on the other hand will be zero if the means are equal and is inversely proportional to the variances. Figure 5 represents these two cases. The assumption of normality can be a critical issue if there is no knowledge of the distributions. Nevertheless, the discrimination power can still be exploited.

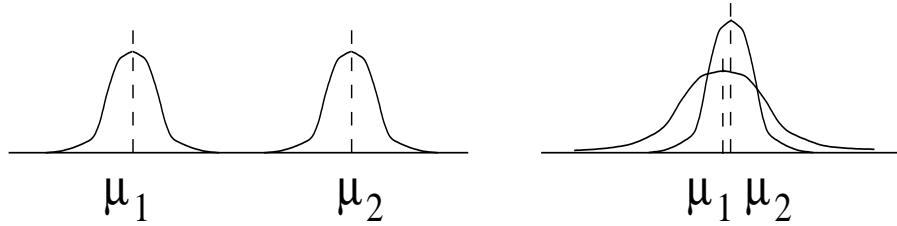


Fig. 5. Bhattacharyya distance cases (a) different means with similar variances (b) Similar means, different variances.

### 3.2 Bhattacharyya Space

The Bhattacharyya space,  $B_{IP}(i, p)$ , is defined as:

$$L_p \times L_i; B_{IP}(i, p) : L_p \times L_i \rightarrow D_B(S_{k_1}^i, S_{k_2}^i). \quad (6)$$

where each class pair,  $p$ , between classes  $k_1, k_2$  at measurement  $i$  will have a Bhattacharyya distance  $D_B(S_{k_1}^i, S_{k_2}^i)$ , and will produce a Bhattacharyya space of dimensions  $N_p = \binom{N_k}{2}$  and  $N_i = 7o : N_p \times N_i$  where  $o$  is the order of the OP and  $N_k$  the number of classes. The domains of the Bhattacharyya space are  $L_i = \{1, 2, \dots, 7o\}$  and  $L_p = \{(1, 2), (1, 3), \dots, (k_1, k_2), \dots, (N_k - 1, N_k)\}$ .

The Bhattacharyya Space is a bivariate state from which two marginal distributions can be extracted:

$$B_I(i) = \sum_{p=1}^{N_p} B_{IP}(i, p) = \sum_{p=1}^{N_p} D_B(S_{k_1}^i, S_{k_2}^i), \quad i = 1, \dots, N_i \quad (7)$$

$$B_P(p) = \sum_{i=1}^{N_i} B_{IP}(i, p) = \sum_{i=1}^{N_i} D_B(S_{k_1}^i, S_{k_2}^i), \quad p = 1, \dots, N_p. \quad (8)$$

The marginal over the class pairs,  $B_I(i)$  sums the Bhattacharyya distance of every pair of a certain feature and thus will indicate how discriminant a certain sub-band OP filter is over the whole combination of class pairs. The marginal  $B_P(p)$  sums the Bhattacharyya distance for a particular pair of classes over the whole measurement space and reveals the discrimination potential of particular pairs of classes when multiple classes are present.

To visualise the previous distribution, the Bhattacharyya Space and its two marginal distributions were obtained for figure 1 (f). Figure 6 shows: (a)  $B_{IP}(i, p)$ , (b)  $B_I(i)$  and (c)  $B_P(p)$ . These graphs yield useful information toward the selection of the features for classification. The most discriminant features for the training data presented are  $S^{19,18,11,\dots}$ . A certain periodicity is revealed in the following dimensions of the measurement space; 1, 7, 14, 21, 28, which have the lowest values (this is clearer in the marginal  $B_I(i)$ ). These measurements correspond to low pass filters of the OP. Since the textures that make up this mosaic have been deliberately histogram equalised, the low pass features provide the lowest discrimination power.

The marginal  $B_P(p)$ , where the index of  $p$  correspond to the pairs  $L_p = \{(1, 2), (1, 3), \dots, (k_1, k_2), \dots, (N_k - 1, N_k)\}$ , can be useful to identify certain pairs of textures which are difficult to segment.

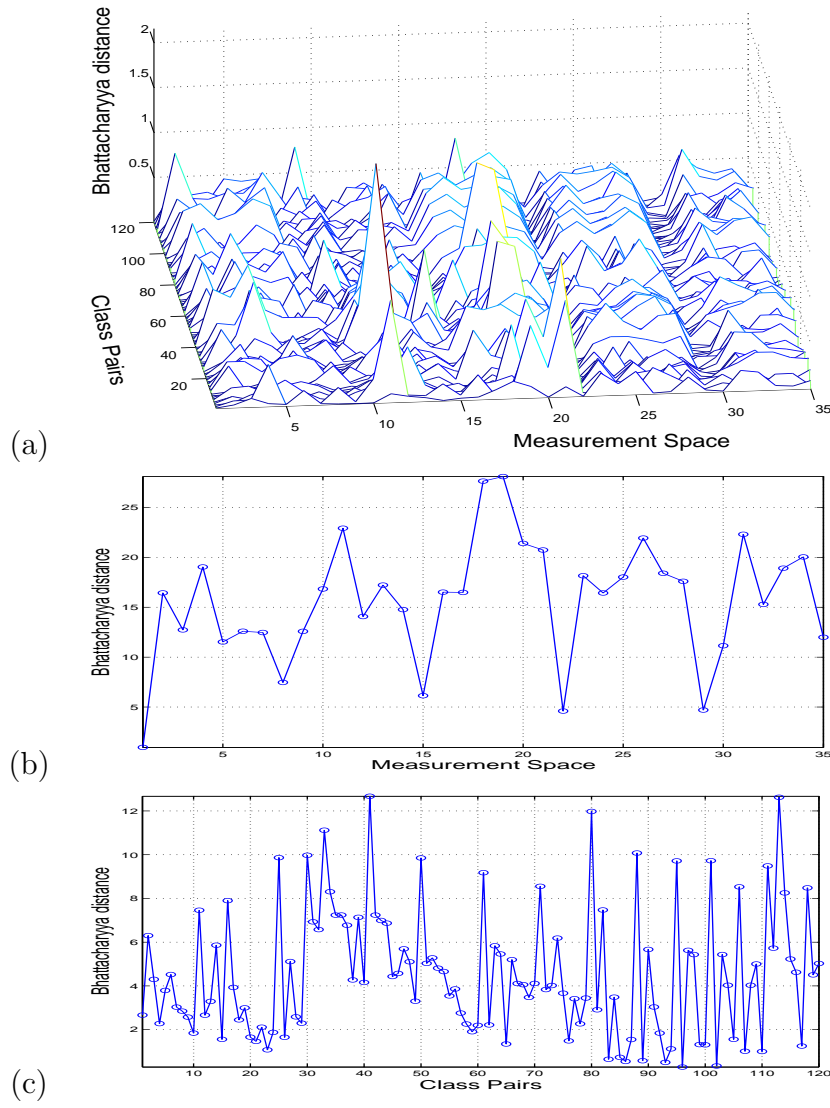


Fig. 6. (a) The Bhattacharyya Space  $B_{IP}(i, p)$  for the Natural Textures image and its corresponding marginals (b)  $B_I(i)$ , (c)  $B_P(p)$ . The index of Class Pairs correspond to the pairs  $L_p = \{(1, 2), (1, 3), \dots (k_1, k_2), \dots (N_k - 1, N_k)\}$ .

### 3.3 Order Statistics for Feature Ranking

If the marginal  $B_I(i) = \{B_I(1), B_I(2), \dots B_I(70)\}$ , is sorted in increasing order, its order statistic will be:

$$\begin{aligned} B_{(I)}(i) &= \{B_{(I)}(1), B_{(I)}(2), \dots B_{(I)}(70)\}, \\ B_{(I)}(1) &\leq B_{(I)}(2) \leq \dots \leq B_{(I)}(70). \end{aligned} \tag{9}$$

where  $^{max}_i(B_I(i)) = B_{(I)}(70)$ ,  $^{min}_i(B_I(i)) = B_{(I)}(1)$  and  $B_I(i) = B_{(I)}(j)$ . The domain  $L_j = \{\dots, j, \dots\}$  provides a particular route for the state space search.

In other words, a re-ordering of the elements of Measurement space  $S^i$  before being sequentially provided to the classifier. The dimensions of the set remain the same as of the measurement space:  $N_j = N_i$ .

Figure 8 exemplifies this for a 4-measurement state space. It is important to mention two aspects of this selection process. First, the Bhattacharyya space is constructed on training data. Second, the individual Bhattacharyya distances are calculated between pairs of classes. As a result of these two aspects, there is no guarantee that the feature selected will improve the classification of the whole data space, they can be mutually redundant or may only improve the classification for a pair of classes but not the overall classification [23].

Thus the conjecture to be tested then is whether the classification can be improved in a *best-first*, sequential selection defined by the Bhattacharyya space order statistics. The natural textures image was classified with several sequential selection strategies:

- Following the unsorted order of the measurement space:  $S^1, S^2, S^3$  etc.
- Following the marginal  $B_{(I)}(i)$  in decreasing order:  $S^{19}, S^{18}, S^{11}$  etc.
- Following the marginal  $B_{(I)}(i)$  in increasing order:  $S^{28}, S^{21}, S^{15}$  etc. (The converse conjecture is that the reverse order should provide the worst path for the classification.)
- Three random permutations.

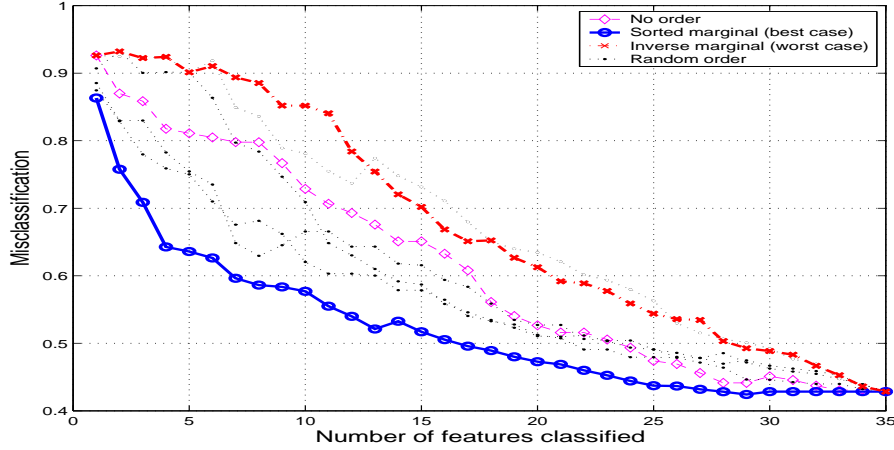


Fig. 7. Misclassification error for the sequential inclusion of features to the classifier for the figure 1 (f).

The sequential misclassification results of the previous strategies are presented in Figure 7 where the advantage of the route provided by the  $B_{(I)}(i)$  can be seen.

Although the Bhattacharyya space appears to be the best result, there are some features that when included increase the misclassification. A heuristic method is proposed to overcome this problem. If the whole state space is



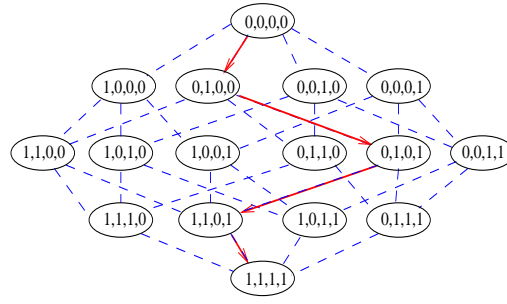


Fig. 8. State Space for sequential selection following the route determined from the Bhattacharyya Space .

traversed up to the state  $\{1, 1, \dots, 1\}$ , a misclassification graph will show which the particular effect to the misclassification (positive/negative) of each feature when included in the classifier. From the graph shown in figure 7 It can be seen that most of the features contribute positively to the classification with the exception of  $B_{(I)}(14, 30)$ , and the last five features  $B_{(I)}(31 - 35)$  leave the classification unchanged. These features can be removed from the classification procedure:

$$S^j \in v \text{ if } \lambda(S^1, S^2, \dots, S^j) > \lambda(S^1, S^2, \dots, S^{j-1}) \quad (10)$$

In the previous example, the set of features to be included in the classifier will be:  $L_m = L_j \setminus \{14, 30 - 35\}$ .  $L_m$  is the domain of the *Feature Space* a reduced and ordered version of the Measurement Space:  $S_F \subset S$ ,  $S^m \in S_F$ ,  $S^m \in S$ ,  $N_m \leq N_i$ . The dimensions of the Feature Space are  $L_r \times L_c \times L_m$ .

Another solution that is provided by the order statistic of the Bhattacharyya Space marginal is the option to select a predetermined number of features as the *reduced set* or sub-space used for classification. This can be use particularly in cases where it can be computationally very expensive just to obtain the whole measurement space. Then, based on the training data, just a few measurements are generated based on the first  $n$  features provided by the Bhattacharyya space.

## 4 Results

Table 2 presents characteristics and classification details for the 9 images. The OP sub-band filtering was used to generate the measurement space of 35 dimensions. This was classified with a single resolution algorithm ( $\hat{a}_k$ ). Then, for each measurement, a *QT* of 5 levels was constructed and the classification was performed at the highest level. Butterfly filters were used to refine the boundaries on the descent of the *QT*. Finally, feature selection was performed

Table 2

Characteristics of the images and their classification results for Single and multiresolution without feature selection and multiresolution with feature selection.

Figure	Size	Classes	Source	Misclassification (%)			
				No selection, 35 Feats		Feature selection	
				Single	Multi	Multi	Features
a	$256 \times 256$	5	Brodatz	9.0	5.2	2.8	23
b	$256 \times 256$	5	MIT	31.7	14.7	14.7	35
c	$256 \times 256$	5	MIT	20.6	22.0	8.4	29
d	$256 \times 256$	5	MIT	20.7	16.1	7.3	14
e	$256 \times 256$	5	MeasTex	17.2	8.5	4.3	20
f	$512 \times 512$	16	Brodatz	32.7	20.4	17.9	23
g	$512 \times 512$	16	MIT	49.5	44.5	32.0	21
h	$256 \times 640$	10	Brodatz	27.9	25.9	14.7	21
i	$256 \times 640$	10	MIT	39.5	32.4	20.2	14
Average				27.6	21.1	13.6	

Table 3

Comparative misclassification (%) results of Malpica [31], Randen [35], Ojala [34] and multiresolution with feature selection. Best results are in **bold**.

Method	Figures									Average
	a	b	c	d	e	f	g	h	i	
Co-occurrence	9.9	27.0	26.1	51.1	35.7	49.6	55.4	35.3	49.1	37.69
Best in Randen	7.2	18.9	20.6	16.8	17.2	34.7	41.7	32.3	27.8	24.13
$p_8$ (Ojala)	7.4	12.8	15.9	18.4	16.6	27.7	33.3	17.6	18.2	18.66
LBP (Ojala)	6.0	18.0	12.1	9.7	11.4	<b>17.0</b>	20.7	22.7	19.4	15.22
Watershed (Malpica)	7.1	<b>10.7</b>	12.4	11.6	14.9	20.0	<b>18.6</b>	<b>12.0</b>	<b>15.3</b>	13.62
Proposed algorithm	<b>2.8</b>	14.8	<b>8.4</b>	<b>7.3</b>	<b>4.3</b>	17.9	32.0	14.7	20.2	<b>13.61</b>

with the Bhattacharyya space and the lowest misclassification was selected. The number of features varied from 14 up to one case (b) in which the 35 features provided the best result.

Two important observations should be made, first, multiresolution classification can improve results over single resolution and second, feature selection can further reduce the misclassification.

To evaluate the performance of the multiresolution classification with feature selection, a comparison was made against the best results of Randen, the results of Ojala [34] who used Local Binary Patterns (LBP) and multidimensional distributions of signed grey-level differences ( $p_8$ ), and those reported by Malpica [31] who used a multichannel watershed-based algorithm with wavelet features. The results of Randen's co-occurrence are included in the comparison since they are widely used.

The final classification results are presented in table 3 and the following observations can be made.

- It should be noted that co-occurrence can easily be outperformed, it is the

worst classification individually and overall.

- The best results presented by Randen were outperformed by all the other methods. Again this was to be expected, since the classification schemes were far more complex than those used by Randen.
- The methods proposed by Ojala outperform those of Randen and have good results, in some cases they are better than Malpica's, but in general they can be outperformed.
- The multichannel watershed-based algorithm (Malpica [31]) presents very good results, in four cases it has the lowest misclassification.
- The multiresolution algorithm with feature selection presents very good results, it is comparable with Malpica's results and in some of the images it provides the best classification.

As an indication of the computational complexity of the algorithm presented, the computation time of the programs running with Matlab version 6.5 R13 running on a Linux platform based on a Pentium 4 CPU 2.80 GHz was measured. The time for the 16-class segmentation of figure 1 (f), was 2.7s for k-means classification at a single resolution and 56.3s for multiresolution with feature selection through the Bhattacharyya space. No systematic attempt to make the code more efficient was made. The classification results are presented below. Figure 9 shows the boundaries on top of the original images, figure 10 shows the results as classified regions, and figure 11 shows the pixels that are correctly classified. These latter results are considered by the author to be the most revealing since showing only the labelled classes or only the boundaries on top of the original images can be misleading.

## 5 Conclusions

A feature selection methodology using a novel *Bhattacharyya Space* has been presented. The Bhattacharyya Space is obtained by calculating the Bhattacharyya distance of pairs of training classes. This method allows the selection of the most discriminant features of a measurement space  $S$  by assessing the class pair or feature marginal of the space. This marginal can be used as a path to follow with a wrapper algorithm. While the solution provided by the Bhattacharyya space is sub-optimal in various ways, when it is combined with a multiresolution classification it can provide the lowest misclassification of the textured images presented by Randen [35].

Another application of the Bhattacharyya space is for detecting which pairs of classes would be particularly hard to discriminate over all the measurement space, and in some cases, the individual use of one point of the space can be also of interest.

The use of the Bhattacharyya Space implies that the number of classes is previously known, thus it is not presented as a method to determine the presence or absence of a number of clusters (one or more) in a certain space. If this is required, other methods like the *Two-point correlation function* or the *distance histogram* proposed by Fatemi-Ghomi [12] could be used.

## 6 Acknowledgments

This work was partly supported by CONACYT - México. Professor Roland Wilson and Dr Nasir Rajpoot are acknowledged for their valuable discussions on this work and help in revising the manuscript. The test data sets along with the software are available on an “as is” basis on the web page: <http://www.dcs.warwick.ac.uk/~creyes/m-vts>.

## 7 Vitae

**Constantino Carlos Reyes-Aldasoro** graduated from UNAM Mexico, (BSEE 1993), Imperial College (MSc Communications and Signal Processing 1994) and Warwick University (PhD Computer Science 2005). Between 1995-2000 he lectured in Mexico. Between 2001-2004 he worked as a GTA at Warwick University and completed a PhD with Dr. Abhir Bhalerao in Medical Image Analysis. He joined The University of Sheffield in 2005 as a Postdoctoral Research Scientist in the Tumour Microcirculation Group with Professor Gill Tozer. His research interests are: Medical Imaging, Microcirculation, Texture, Image Processing and Pattern Recognition.

**Abhir Bhalerao** graduated with a PhD in Computer Science in 1992 (Warwick). Between 1993-1997 he was a senior clinical scientist at Kings Medical School and Guy’s and St. Thomas Hospitals in London and spent 2 years as a Research Fellow at Harvard at the Surgical Planning Laboratory working on artefact correction in MRI data, vessel segmentation and visualisation for surgical intervention. He joined Computer Science at Warwick as faculty in 1998 and has research interests in texture analysis, pattern recognition and medical image processing and has published over 30 refereed articles in these areas.

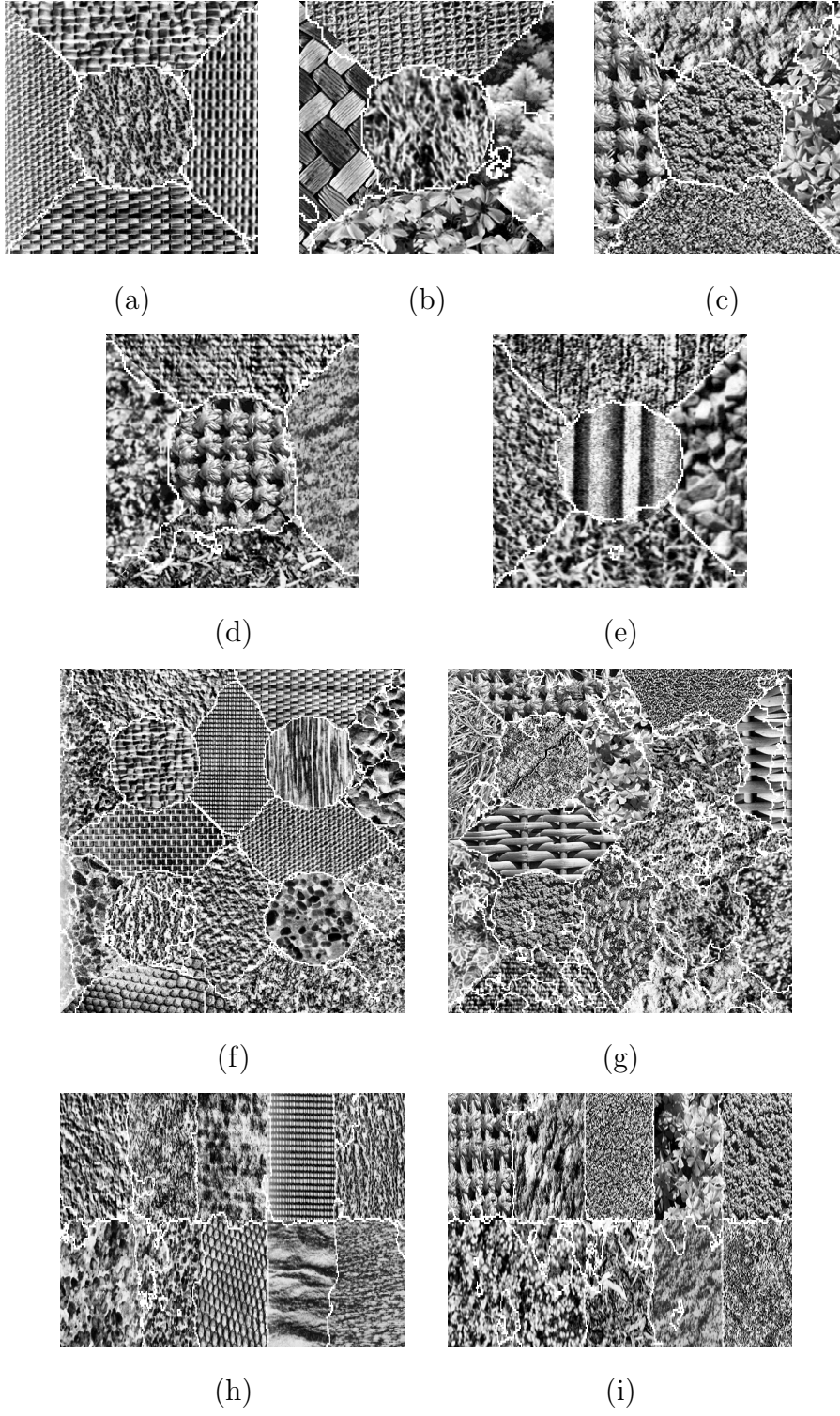


Fig. 9. Classification of the images in figure 1. Classes boundaries are super-imposed in the images.

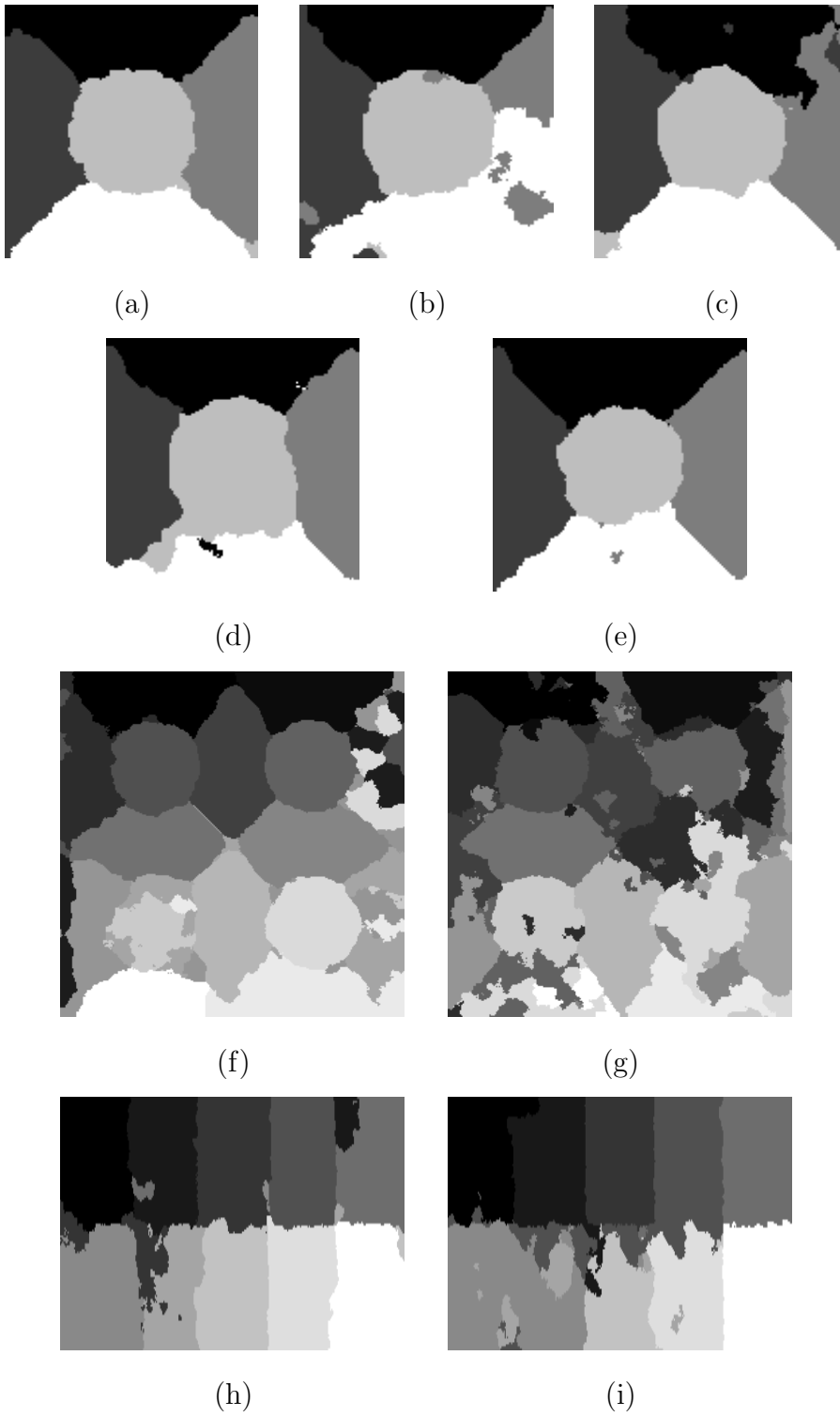


Fig. 10. Classification of the images in figure 1. Classes are presented as different levels of grey.

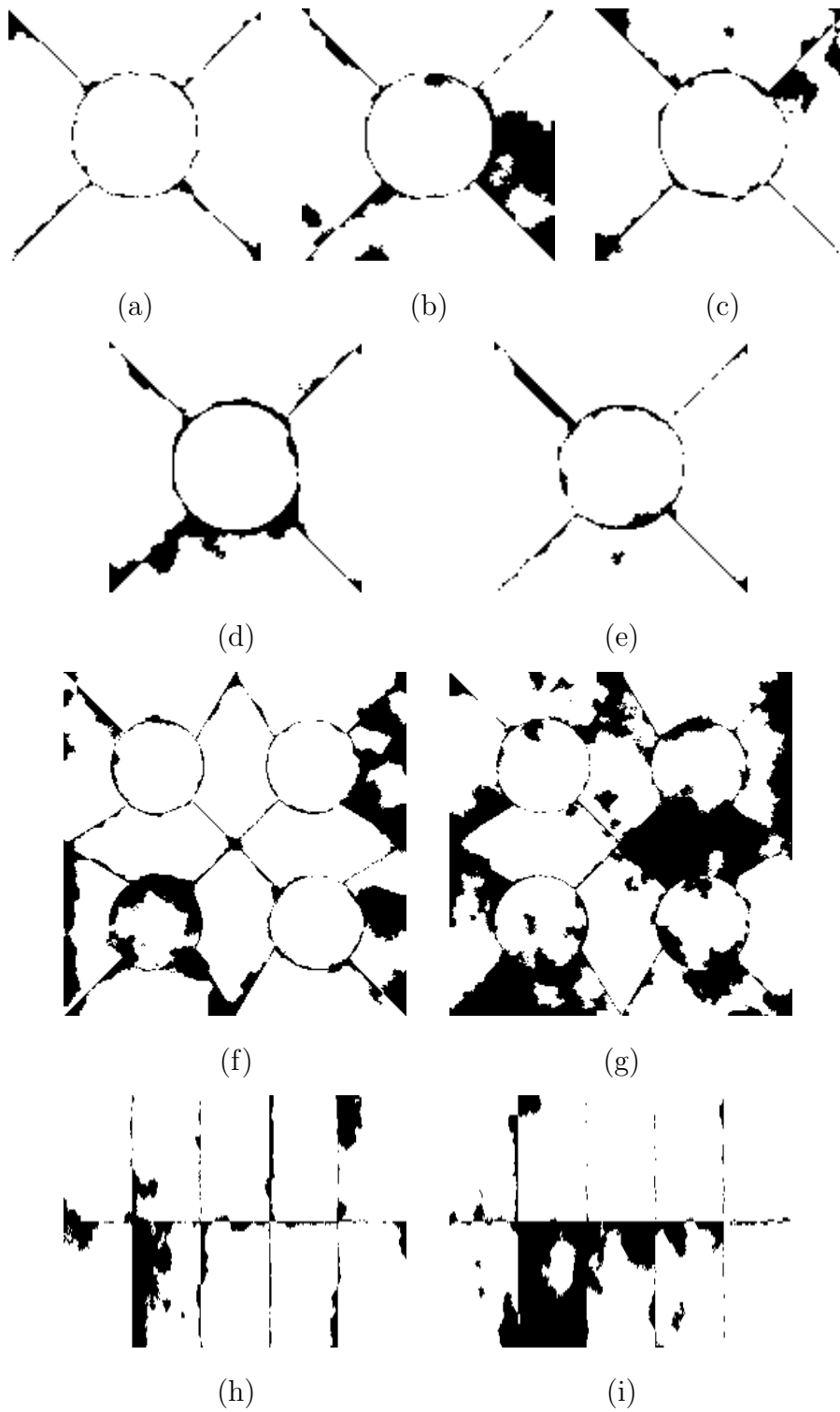


Fig. 11. Classification of the images in figure 1. Pixels that are correctly classified appear in white.

## References

- [1] A. Bhalerao and N. Rajpoot. Selecting Discriminant Subbands for Texture Classification. In *BMVC*, Norwich, UK, September 2003.
- [2] L. Blot and R. Zwiggelaar. Synthesis and Analysis of Solid Texture: Application in Medical Imaging. In *Texture 2002: 2nd Int Workshop on Texture Analysis and Synthesis*, pages 9–14, Copenhagen, 1 June 2002.
- [3] C. Bouman and B. Liu. Multiple Resolution Segmentation of Textured Images. *IEEE Trans. Pattern Anal. Machine Intel.*, 13(2):99–113, 1991.
- [4] P. Brodatz. *Textures: A photographic album for artists and designers*. Dover, New York, U.S.A, 1996.
- [5] P. J. Burt and E. H. Adelson. The Laplacian Pyramid as a compact Image Code. *IEEE Trans. Commun.*, 31(4):532–540, 1983.
- [6] A. Carrillat, T. Randen, L. Sønneland, and G. Elvebakk. Seismic stratigraphic mapping of carbonate mounds using 3D texture attributes. In *Extended Abstracts, Annual Meeting, European Association of Geoscientists and Engineers*, Florence, Italy, May 27-30 2002.
- [7] R. Chellapa and A. Jain. *Markov Random Fields*. Academic Press, Boston, 1993.
- [8] G. B. Coleman and H. C. Andrews. Image Segmentation by Clustering. *Proceedings of the IEEE*, 67(5):773–785, 1979.
- [9] G. R. Cross and A. K. Jain. Markov Random Field Texture Models. *IEEE Trans. Pattern Anal. Machine Intel.*, 5(1):25–39, 1983.
- [10] M. Dong and R. Kothari. Feature subset selection using a new definition of classifiability. *Patt Rec Letts*, 24(9-10):1215–1225, 2003.
- [11] E. R. Dougherty and M. Brun. A probabilistic theory of clustering. *Patt Rec*, 37(5):917–925, 2004.
- [12] N. Fatemi-Ghomi. *Performance measures for wavelet-based segmentation algorithms*. PhD thesis, Centre for Vision, Speech and Signal Processing, University of Surrey, 1997.
- [13] K. Fukunaga. *Introduction to Statistical Patt Rec*. Academic Press, 1972.
- [14] V. Gaede and O. Günther. Multidimensional access methods. *ACM Computing Surveys*, 30(2):170–231, 1998.
- [15] I. Guyon and A. Elisseeff. An Introduction to Variable and Feature Selection. *J of Machine Learning Research*, 3(7-8):1157–1182, 2003.
- [16] D. J. Hand. *Discrimination and classification*. Wiley, Chichester, 1981.



- [17] R. M. Haralick. Statistical and Structural Approaches to Texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.
- [18] R. M. Haralick, K. Shanmugam, and I. Dinstein. Textural Features for Image Classification. *IEEE Trans. Syst., Man, Cybern.*, 3(6):610–621, 1973.
- [19] E. A. Hoffman, J. M. Reinhardt, M. Sonka, B. A. Simon, J. Guo, O. Saba, D. Chon, S. Samrah, H. Shikata, J. Tschirren, K. Palagyi, K. C. Beck, and G. McLennan. Characterization of the Interstitial Lung Diseases via Density-Based and Texture-Based Analysis of Computed Tomography Images of Lung Structure and Function. *Academic Radiology*, 10(10):1104–1118, 2003.
- [20] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Patt Rec*, 24(12):1167– 1186, 1991.
- [21] A. Kadyrov, A. Talepbour, and M. Petrou. Texture Classification with Thousands of Features. In *BMVC*, pages 656–665, Cardiff, UK, 2-5 September 2002.
- [22] T. Kailath. The Divergence and Bhattacharyya Distance Measures in Signal Selection. *IEEE Trans. Commun. Technol.*, 15(1):52–60, 1967.
- [23] J. Kittler. Feature Selection and Extraction. In Y. Fu, editor, *Handbook of Pattern Recognition and Image Processing*, pages 59–83, New York, 1986. Academic Press.
- [24] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.
- [25] V. A. Kovalev, M. Petrou, and Y. S. Bondar. Texture Anisotropy in 3D Images. *IEEE Trans. Image Processing*, 8(3):346–360, 1999.
- [26] R. Kumar, V. K. Jayaraman, and B. D. Kulkarni. An svm classifier incorporating simultaneous noise reduction and feature selection: illustrative case examples. *Patt. Rec.*, 38:41–49, 2005.
- [27] R. Lerski, K. Straughan, L. R. Schad, D. Boyce, S. Bluml, and I. Zuna. MR Image Texture Analysis - An Approach to tissue Characterization. *Mag Res Imag*, 11(6):873–887, 1993.
- [28] S. Li, J. T. Kwok, H. Azu, and Y. Wang. Texture classification using the support vector machines. *Patt. Rec.*, 36:2883–2893, 2003.
- [29] H. Liu and H. Motoda. *Feature Extraction Construction and Selection. A Data Mining Perspective*. Kluwer Academic Publisher, Norwell MA, 1998.
- [30] X. Liu and D. Wang. Appearance-Based Recognition Using Perceptual Components. In *Int Joint Conf in Neural Networks*, pages 1943–1948, Washington, DC, March 2001.
- [31] N. Malpica, J. E. Ortuño, and A. Santos. A multichannel watershed-based algorithm for supervised texture segmentation. *Patt Rec Letts*, 24(9):1545–1554, 2003.

- [32] MIT Media Lab. Vision texture. <http://vismod.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>, 1995.
- [33] T. Ojala, M. Pietikäinen, and D. Harwood. A Comparative Study of Texture Measures with Classification based on Feature Distributions. *Patt Rec*, 29(1):51–59, 1996.
- [34] T. Ojala, K. Valkealahti, E. Oja, and M. Pietikäinen. Texture discrimination with multidimensional distributions of signed gray level differences. *Patt Rec*, 34(3):727–739, 2001.
- [35] T. Randen and J. H. Husøy. Filtering for Texture Classification: A Comparative Study. *IEEE Trans. Pattern Anal. Machine Intel.*, 21(4):291–310, 1999.
- [36] T. Randen, E. Monsen, A. Abrahamsen, J. O. Hansen, J. Schlaf, and L. Sønneland. Three-dimensional texture attributes for seismic data analysis. In *Ann. Int. Mtg., Soc. Expl. Geophys., Exp. Abstr.*, Calgary, Canada, August 2000.
- [37] T. Reed and J. du Buf. A Review of Recent Texture Segmentation and Feature Extraction Techniques. *CVGIP: Image Understanding*, 57(3):359–372, May 1993.
- [38] C. C. Reyes-Aldasoro. *Multiresolution Volumetric Texture Segmentation*. PhD thesis, Department of Computer Science, University of Warwick, April 2005.
- [39] C. C. Reyes-Aldasoro and A. Bhalerao. Volumetric Texture Description and Discriminant Feature Selection for MRI. In C. Taylor and A. Noble, editors, *Proceedings of Information Processing in Medical Imaging*, pages 282–293, Ambleside, UK, July 2003.
- [40] H. Samet. The Quadtree and Related Hierarchical Data Structures. *Computing Surveys*, 16(2):187–260, 1984.
- [41] L. R. Schad, S. Bluml, and I. Zuna. MR Tissue Characterization of Intracranial Tumors by means of Texture Analysis. *Mag Res Imag*, 11(6):889–896, 1993.
- [42] P. Schroeter and J. Bigun. Hierarchical Image Segmentation by Multi-dimensional Clustering and Orientation-Adaptive Boundary Refinement. *Patt Rec*, 28(5):695–709, 1995.
- [43] M. Segovia-Martínez, M. Petrou, V. A. Kovalev, and P. Perner. Quantifying Level of Brain Atrophy Using Texture Anisotropy in CT Data. In *Medical Imaging Understanding and Analysis*, pages 173–176, Oxford, UK, July 1999.
- [44] G. Smith. Meastex image texture database and test suite. <http://www.cssip.uq.edu/meastex/meastex.html>, 1997.
- [45] M. Spann and R. Wilson. A quad-tree approach to image segmentation which combines statistical and spatial information. *Patt Rec*, 18(3/4):257–269, 1985.
- [46] C. Tai and K. Baba-Kishi. Microtexture studies of PST and PZT Ceramics and PZT Thin Film by Electron Backscatter Diffraction Patterns. *Textures and Microstructures*, 35(2):71–86, 2002.

- [47] M. Tuceryan and A. K. Jain. Texture Analysis. In C. H. Chen, L. F. Pau, and P. S. P. Wang, editors, *Handbook of Pattern Recognition and Computer Vision*, pages 207–248. World Scientific Publishing, 1998.
- [48] M. Unser. Texture Classification and Segmentation Using Wavelet Frames. *IEEE Trans. Image Processing*, 4(11):1549–1560, 1995.
- [49] M. Unser and M. Eden. Multiresolution Feature Extraction and Selection for Texture Segmentation. *IEEE Trans. Pattern Anal. Machine Intel.*, 11(7):717–728, 1989.
- [50] J. Weszka, C. Dyer, and A. Rosenfeld. A comparative Study of Texture Measures for Terrain Classification. *IEEE Trans. Syst., Man, Cybern.*, 6(4):269–285, April 1976.
- [51] R. Wilson and C.-T. Li. A Class of Discrete Multiresolution Random Fields and its Application to Image Segmentation. *IEEE Trans. Pattern Anal. Machine Intel.*, 25(1):42–56, 2003.
- [52] R. Wilson and M. Spann. *Image Segmentation and Uncertainty*. John Wiley and Sons Inc., New York, 1988.
- [53] Y. Zhan and D. Shen. Automated Segmentation of 3D US Prostate Images Using Statistical Texture-Based Matching Method. In *MICCAI*, pages 688–696, Canada, November 16-18 2003.