A Generative Probabilistic Oriented Wavelet Model for Texture Segmentation

Inna Stainvas*and David Lowe
Neural Computing Research Group
Information Engineering
Aston University
United Kingdom
stainvai,d.lowe@aston.ac.uk

17 October 2001; Revised 12 August 2002

*Corresponding author. Address: NCRG, Information Engineering, Aston University, Aston street, Birmingham B4 7ET, United Kingdom, Phone +44 (0)121 359 3611 ext. 4244; Fax +44 (0)121 333 4586

Abstract

This paper addresses image segmentation via a generative model approach. A Bayesian network (BNT) in the space of dyadic wavelet transform coefficients is introduced to model texture images. The model is similar to a Hidden Markov model (HMM), but with non-stationary transitive conditional probability distributions. It is composed of discrete hidden variables and observable Gaussian outputs for wavelet coefficients. In particular, the Gabor wavelet transform is considered.

The introduced model is compared with the simplest joint Gaussian probabilistic model for Gabor wavelet coefficients for several textures from the Brodatz album [1]. The comparison is based on cross-validation and includes probabilistic model ensembles instead of single models. In addition, the robustness of the models to cope with additive Gaussian noise is investigated. We further study the feasibility of the introduced generative model for image segmentation in the novelty detection framework [2]. Two examples are considered: (i) sea surface pollution detection from intensity images and (ii) image segmentation of the still images with varying illumination across the scene.

Keywords: Dyadic wavelet transform, Gabor wavelet transform, generative probabilistic model, Bayesian networks and ensembles, image texture segmentation, novelty detection.

Glossary:

```
BNT  - Bayesian network
HMM  - Hidden Markov model
WT   - Wavelet transform
WTC  - Wavelet transform coefficients
GWT  - Gabor wavelet transform
GWTC - Gabor wavelet transform coefficients
MAP  - Maximum a posterior principle
ML   - Maximum likelihood
EM   - Expectation-Maximization
MRF  - Markov random field
HMTM - Hidden Markov tree model
CPD  - Conditional probability distribution
CPT  - Conditional probability table
SNR  - Signal to noise ratio
CV   - Cross-validation
PDF  - Probability density function
GMM  - Gaussian mixture model
```

Introduction

Image segmentation is a difficult and yet very important problem arising in many visual applications such as medical imaging, automated monitoring,

document processing, remote sensing and many others. The main goal of image segmentation is to decompose an image into its constituent parts or objects [3, 4]. This becomes possible since real objects have homogeneous physical properties that should be reflected in images. In some cases, an assumption of the homogeneity of the object gray level intensity or color is sufficient and works well in practice. In many others, this assumption is violated, and instead image objects are assumed to be represented as repeated patterns called visual *textures* (see Figures 2a,d for some texture examples). Though textures are easily recognized by humans, there is no a unique and strict mathematical definition of the latter [5, see for review and other definitions].

Segmentation based on texture properties is referred to as *texture segmentation*. The level to which segmentation is carried and the approaches to address the problem crucially depend on the particular application, i.e. its aims and the available information. In general, the less is known about the possible number of objects (textures) in the image scene and their appearance, the more difficult the problem becomes. In some applications, the image parts should be additionally classified into certain texture categories, such as soil, sand, grass, etc. This type of the segmentation problem is referred to as a *texture classification* problem and it assumes that representatives of all possible textures that may appear in the image are available. There is also a constrained variant of this problem when only the object (texture) of interest should be found in the image and only the information about this particular texture or its antipode appearance is available beforehand; we refer to this type of the segmentation problem as *texture detection*.

Texture classification, detection and image segmentation, in general, are difficult representatives of statistical pattern recognition when the data is high-dimensional. In the framework of statistical pattern recognition, texture (image) classification and detection emerge as supervised and semi-supervised image pixel classification tasks, respectively. In the most difficult case of the texture (image) segmentation when no a priori information is available to us, texture segmentation emerges as unsupervised image pixel classification task.

An important element of the supervised/unsupervised classification using a Bayesian approach, leading to a MAP (maximum a posterior principle) when misclassification loss functions are the same for all classes [6], is estimating posterior class probabilities $p(c = i|x)$ given the observation $x \in \mathcal{R}^n$. It is in general, a difficult task and one of the ways to avoid a direct estimation of the posterior class probabilities is using classifiers based on generative models.

Classifiers based on generative models estimate posterior class probabil-

ities using the Bayes' rule:

$$p(c = i|x) = \frac{p(x|c = i)p(c = i)}{\sum_{i=1}^{M} p(x|c = i)p(c = i)},$$

where M is a number of classes. If all the class priors are the same, estimation of the posterior probabilities is completely replaced by estimating conditional class probabilities $p(x|c = i)$. Though this problem may also be quite difficult, it may still be easier than the original one. Moreover, since our main goals are classification, detection and segmentation, less complex models and less data than for a texture synthesis problem[1], may be required. The generative model classifiers may be easily updated when a new class is added or removed, or when new features independent from the previous are introduced. Using generative models for texture classification implies that *textures are samples of ergodic stochastic processes* [7].

One can easily recognize MRFs (Markov Random Fields) [8, 9] as a type of generative model classifier. Though MRFs can describe a wide range of image distributions, they are known to be computationally intensive. The convergence and computation of MRFs grow exponentially with the clique size, that make them unrealistic in practice.

Recently, with the rapid development of wavelet techniques an interest in the generative model classifier approaches was revived. New approaches based on modeling the texture images in the wavelet domain have been proposed [10, 11, 12]. Due to multi-resolution properties of the wavelet transform (WT) and locality of wavelet basis functions, a wavelet image representation turns out to be simple, i.e. it is sparse [13, 14] and it has small redundancy. This allows the encoding of WT coefficients by simple probabilistic models that are factorized across scale. The models can be trained from the single image due to the ergodicity assumption, that sampling over an image space domain and over a random field are equivalent [7, 15].

Despite, the common assumptions and ideas underlying these approaches, they still differ in the type of the wavelet transform used and probabilistic models imposed. The multi-scale statistical model [11] assumes a wavelet transform with a pyramidal structure but with a trivial non-parametric conditional distributions as a ratio of Parzen window density estimators. Alternatively Hidden Markov Tree Models (HMTMs) [10] are parametric models with discrete hidden states and Gaussian observable variables introduced for separable 2D wavelet transforms. In addition, these models are limited by the independent band assumption, i.e. wavelet coefficients corresponding to three different orientations: horizontal, vertical and diagonal, are assumed to be independent. Finally, random cascades on wavelet trees [12] have been

---

[1]The problem of generating textures as samples from the probabilistic model.

introduced for the steerable pyramid[2] [16] and the pyramid (wavelet) coefficients are described as Gaussian scale mixtures with the continuous (hidden) scale variables obeying a multi-scale autoregressive process. This model was used for image denoising and wavelet coefficient coding.

In this paper, we continue this line of thought and introduce a generative probabilistic model for Gabor wavelet transform (GWT) coefficients to solve two types of image segmentation problems: *texture classification and texture detection*. The GWT has been successfully and extensively used for texture analysis [17, 18, 19]; it is shift-invariant, has optimal spatial/frequency localization properties, has finer orientation selectivity than a separable realization of the WT [13] and is biologically motivated. The GWT (Sect. 1) may be more efficient for the analysis of complex oriented textures than the 2D separable WT.

Though the model is demonstrated using a Gabor WT, it is quite general and may be easily generalized to probabilistic modeling of the output coefficients of any bank of filters depending on scale and orientation parameters, such as the dyadic oriented wavelet transform [13] or a bank of difference of Gaussian filters [20] considered at different scales and orientations.

The introduced generative model (Sect. 2) has the form of a Bayesian network (BNT) [21] and is similar to HMTMs as it has a mixed discrete hidden state and continuous observation variable model. However, the independent band assumption is relaxed and replaced by the wavelet coefficient independence within the scale. Similar to [12] the model encodes orientation and scale dependencies simultaneously. In contrast to the reviewed works, the dyadic WT is used instead of the decimated (pyramidal) WT and our main goal is a texture segmentation instead of texture synthesis. The Gabor BNT parameters are found using the EM (expectation maximization algorithm) and image classification is based on the Bayesian classifier that is equivalent to classification by ML (maximum likelihood) in our case[3].

In order to perform texture detection, the novelty detection approach [2] is used. In the texture detection problem, probabilistic model for only one texture of interest is learned; and final detection is based on identifying the image pixels as belonging to the texture, if they get sufficiently large likelihood value under the learned probabilistic model.

The introduced model is compared (Sect. 3) with the simplest joint Gaussian probabilistic model for Gabor wavelet coefficients for several textures from the Brodatz album [1]. The comparison is based on cross-validation and includes probabilistic model ensembles instead of single models. In addition, the robustness of the models to cope with additive Gaussian noise is investigated. We further study the feasibility of the introduced generative

---

[2]The 2D separable wavelet transform may be considered as a partial case of the steerable pyramid

[3]class priors are assumed to be the same

5

model (Sect. 4) for (i) sea surface pollution detection from intensity images and (ii) object detection from still images with varying illumination across the scene.

1 Gabor Wavelet Transform

Application of the oriented WT has been motivated by many physiological experiments discovering a class of cells in the mammalian visual cortex, whose responses depend on the frequency and orientation of the visual stimuli [13]. In particular, it has been shown [17] that these impulse responses can be approximated by Gaussian windows modulated by a harmonic wave. These observations motivated a wide use of the Gabor WT in the computer vision study.

The Gabor WT is a type of dyadic oriented wavelet transform [13] with complex valued filters defined by:

$$g(\vec{r}) = C_\sigma exp(-\frac{\parallel \vec{r} \parallel^2}{2\sigma^2}) exp(2\pi i \vec{k} \cdot \vec{r}), \quad C_\sigma = \frac{1}{2\pi\sigma^2}, \tag{1}$$

where $\vec{r} = (x, y)$ and $\vec{k} = (f\cos\theta, f\sin\theta)$ is a vector defining a radial frequency f and orientation θ of the cosinusoidal/sinusoidal modulation waves. The MTF (modulated transfer function) of this filter is given by:

$$\hat{g}(\vec{w}) = exp(-\frac{\parallel \vec{w} - \vec{k} \parallel^2}{2\sigma^{\star 2}}), \quad \sigma^\star = \frac{1}{2\pi\sigma}. \tag{2}$$

Equation (2) shows that in the frequency domain the energy of the filter is concentrated at the frequency $\vec{w} = \vec{k}$ and its effective support is inversely proportional to the scale parameter σ.

The filter parameters are sampled by the equal logarithmic frequency band scheme; i.e. Gaussian filters are distributed over the frequency domain in such a way that their size increases by a factor 2 and adjacent Gaussians intersect at positions where their respective magnitudes have values of half of their maxima. Such sampling guarantees almost a complete covering of the frequency domain [13]. As a result one gets the following scheme:

$$f_j = 2^{j-1}f_0, \quad \sigma_j^\star = \frac{f_j}{3\sqrt{2ln2}}, \quad j = 1, 2 \ldots log_2(N/2). \tag{3}$$

The initial frequency is equal to $f_0 = \frac{\sqrt{2}}{N}$, where N is the texture sample size in pixels. Parameter j stands for an octave number. For some textures, the lowest radial frequencies are not very useful, because the corresponding features are too coarse. So actually, one can start from a larger octave number $j_0 > 1$ and compensate low frequencies with a Gaussian filter with the MTF given by $exp(-\frac{\parallel \vec{w} \parallel^2}{2\sigma_{j_0-1}^{\star 2}})$. The orientation parameter θ is sampled
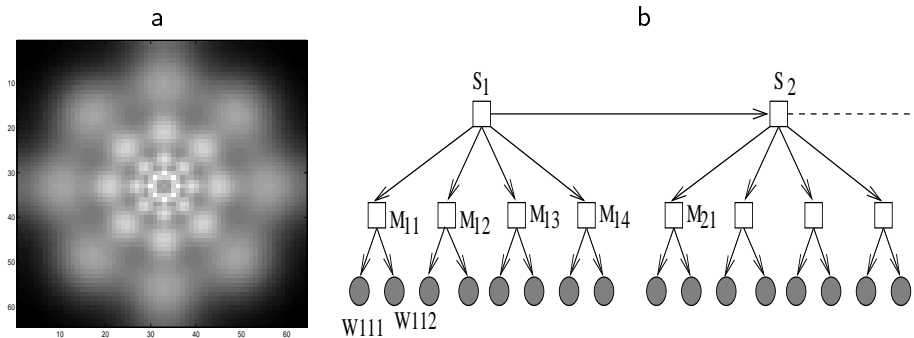
Figure 1: **Gabor WT and Gabor BNT:** a. Sum of the MTFs of the even Gabor WT for $j = 2 \ldots 5$ with four orientations and a low-pass Gaussian filter; $N = 64$; b. Gabor BNT. Note how the BNT for the Gabor WTCs (Figure 1b) exploits the rosette like structure of the Gabor WT filters in the frequency domain (see Figure 1a).

with $\Delta\theta = \pi/K$, where K is a number of considered orientations starting from $\theta_0 = 0$, so that $\theta_k = (k-1)\Delta\theta$. Filter (1) is a complex filter, consisting of a pair of real filters corresponding to the real g_R and imaginary g_I parts. In applications often only real-valued even symmetric Gabor filters corresponding to the real part of the filter (1): $g_R = C_\sigma exp(-\frac{\|\vec{r}\|^2}{2\sigma^2}) \cos(2\pi\vec{k} \cdot \vec{r})$ is used. Using the even-symmetric filters only has been justified by psychophysical studies [18]; moreover, it speeds up the learning and segmentation/classification processes[4]. We use a fast separable realization of the Gabor WT in the spatial domain as was proposed in [19]. GWT coefficients are obtained convolving an image x with the family of filters (1) sampled by scheme (3):

$$W_{jk1} = x * g_R(f_j, \theta_k), \quad W_{jk2} = x * g_I(f_j, \theta_k). \tag{4}$$

The sum of the MTFs of the low-pass filter and the even part of the GWT for $j = 2 \ldots 5$ and $K = 4$ orientations for the case of $N = 64$ pixels is shown in Figure 1a and an example of the GWT coefficients in Figures 7c-d.

2   Gabor WT coefficient modeling

The Gabor WT obeys similar *primary and secondary* properties as the 2D separable WT [22]. Primary properties include: *locality* of the filters in the spatial and frequency domain, *multi-resolution and compression*. The

---

[4]our code is written and algorithm given for a general case of complex GWT coefficients, however only the even part is used in simulations presented.

latter means that the WT leads to a sparse representation, i.e. there are many small coefficients. The primary properties have been used as an argument to the approximate independence of the WT coefficients (WTCs) and as an assumption of the sharply peaked exponential distribution of the WTCs [13, 14]. However, the WT can not completely decorrelate real-world signals or images; these residual dependencies always remain and are referred to as secondary properties. The secondary properties mean that large/small coefficients tend to propagate across scales and, *additionally*, across adjacent orientations in the context of the Gabor WT. We propose to model the primary and secondary properties of the Gabor WT coefficients by the Bayesian network [21, 23, see for introduction on graphical models] depicted in Figure 1b.

This network is similar to a Hidden Markov model [24], but with non-stationary transitive conditional probability distributions (CPDs). Similar to a way the HMTM in the wavelet domain inherits the tree structure of the WT, the BNT for the Gabor WTCs (Figure 1b) exploits the rosette like structure of the Gabor WT filters in the frequency domain (see Figure 1a). This network consists of the discrete and Gaussian continuous nodes. Nodes S_j and M_{jk} are discrete and hidden and nodes W_{jki} are continuous, observable and correspond to GWTC (4). Node S_j corresponds to an octave number j and represents the overall scale activity. We assume that the discrete variable S_j is binary; when its value is zero ($S_j = 0$) the scale is non-active and when $S_j = 1$, it is in an active state. Discrete variables M_{jk}, $k = 1 \ldots K$ correspond to K orientations of the Gabor WT for the scale σ_j and have a similar activity state interpretation as S_j.

From the Markovian properties of the Gabor BNT graph, one can see that the Gabor WTCs corresponding to different scales are assumed to be conditionally independent given a corresponding parent scale variable S. Similarly, the Gabor WTCs are assumed to be conditionally independent within the scale given the values of the corresponding parent orientation variable M. This network imposing conditional independence on the WTCs connects them (introduces dependencies) through the hidden state variables. The joint probability of the complete variable $(\mathbf{W}, \mathbf{S}, \mathbf{M})$ is read from the BNT graph as:

$$p(\mathbf{W}, \mathbf{S}, \mathbf{M}) = \prod_{i=1}^{2} \prod_{k=1}^{K} \prod_{j=j_0}^{J} P(W_{jki}|M_{jk})P(M_{jk}|S_j)P(S_j|S_{j-1}), \qquad (5)$$

where $P(S_{j_0}|S_{j_0-1}) \equiv P(S_{j_0})$ stands for a prior probability of the binary variable S_{j_0} (j_0 is a number of the band octave from which the modeling starts).

When only the even (cosine) WTCs are considered each node M_{jk} has only one child wavelet node W_{jk1} and the product over i is dropped in equation (5). If desired, a coefficient corresponding to a low-pass filter may

also be considered. It is assumed to be independent of the WTCs, thus, its PDF is modeled separately and multiplied by $p(\mathbf{W}, \mathbf{S}, \mathbf{M})$.

Gabor BNT parameters include conditional probability tables (CPTs) for discrete variables S_j, M_{jk} and means and variances for W_{jki}. Therefore, the following parameters have to be defined: (i) prior probability $p(S_{j_0} = 0)$, (one parameter); (ii) transitive probabilities between scale activity state variables, $p(S_{j+1} = 0|S_j = l)$, where $l \in \{0, 1\}$ and $j = j_0 \ldots J - 1$ ($2(J - j_0)$ parameters); (iii) transitive probabilities between scale and orientation activity state variables, $p(M_{jk} = 0|S_j = l)$ ($2K(J - j_0 + 1)$ parameters); (iv) conditional means $\mu_{jki,l}$ and variances $\sigma_{jki,l}$ of the wavelet coefficients given that the parent orientation node M_{jk} is in the state $l \in \{0, 1\}$ ($8K(J - j_0 + 1)$ parameters for even and odd parts of WTCs). Since the Gabor WT filters are band-width filters, theoretical means are approximately equal to zero[5]. In addition, we allow tying CPTs between scale and orientation parameters, so that $p(M_{jk} = 0|S_j = l)$ do not depend on j. This significantly reduces the number of parameters and in addition, imposes new scale/orientation dependencies on the WTCs.

The joint WTC probability is a mixture of Gaussians with diagonal matrices and with a number of mixture components equal to $2^{K(J-j_0+1)}$, but with the mixture coefficients constrained by the Gabor BNT structure. A direct modeling of a Gaussian mixture with the same number of mixture components, requires imposing an appropriate prior on the mixture coefficients and covariance matrices and is not simple. The marginal distribution (non-joint) of WTCs is a mixture of two Gaussians that approximates exponential distributions quite well; this is due to modeling small coefficients by a Gaussian with a small σ and large coefficients with σ large.

We also consider a jointly Gaussian distribution model in the Gaborian space as a baseline for comparison. This model can efficiently capture linear correlations between wavelet coefficients and thus is a good model for comparison. We also note that since the Gabor WT basis is not a strictly complete basis (as the lowest/highest pass filters are neglected), data modeling by a joint Gaussian distribution in the original signal space and Gaborian space are not equivalent. The network parameters are found using the EM algorithm that is straightforward for the considered Bayesian network and is similar to the Baum-Welsh (Forward-backward) algorithm used in HMM [25, 23, 24].

---

[5]This is due to the ergodicity assumption: $E[W] \approx \int x * g \, ds = \mathcal{F}[x * g]|_{w=0} = \mathcal{F}[x]|_{w=0}\mathcal{F}[g]|_{w=0}$, where s is a spatial variable, x is a signal, g is a band-pass filter with $\mathcal{F}[g]|_{w=0} \approx 0$ and $\mathcal{F}$ is a Fourier transform.

3 Texture classification

For performance evaluation of generative Gabor BNTs for texture classification, images from the Brodatz album [1] have been used. The Brodatz data consists of 112 monochrome images of different textures of size 512×512 pixels and is available on-line[6]. Textures are referred to by the number in parenthesis (i.e. D12) that corresponds to the page number in the Brodatz texture book [1]. Despite a wide use of the Brodatz data, there is not a single benchmark technique comparing different algorithms, since the latter are applied in different scenarios and use different measures and data blocks. We choose nine images from the Brodatz album [1]; this choice was constrained to such images that any sub-image of size 64×64 pixels cropped from the entire image is sufficient for perceptual discrimination. This means that the chosen textures are regular and are likely to satisfy the ergodicity condition[7].

First the GWT has been applied to each image and then obtained wavelet coefficients (as images) have been sampled uniformly with the rate of 64 pixels per row and column to get texture samples in the wavelet domain. These samples have been disjointly split into $S = 10$ cross-folds, in order to estimate misclassification error using cross-validation (CV) [26]. Several schemes to use CV may be proposed in the context of generative classifiers. The one that has been used is described below.

Let us enumerate different textures by an index $r = 1, 2, \ldots, M$, i.e. instead of saying that the texture represents grass, sand, soil or whatever it is referred by its number. Now, let f_{rs} be the sth cross-fold for the rth-texture and $p_{rs}(\mathbf{W})$ be a probability assigned to a sample $\mathbf{W}$ by the Gabor BNT of the rth-texture with the parameters trained on all the texture samples except a cross-fold f_{rs}. In order to classify a sample $\mathbf{W} \in f_{rs}$ in the Gaborian space, the ML classifier should compare probability $p_{rs}(\mathbf{W})$ with the probabilities assigned by the Gabor BNTs corresponding to textures $\rho \neq r$. But there are S probability models for each ρth-texture and none of them have seen training data for texture r during training; i.e. there exist S^{M-1} (M is a number of textures) different classifiers to estimate error on the cross-fold f_{rs}. In order to avoid this computational burden and stabilize results, we propose to average $p_{\rho s}(\mathbf{W})$ over cross-folds per each alternative texture[8], to get a simple *texture ensemble* $p_\rho^e(W) = \frac{1}{S} \sum_{s=1}^{S} p_{\rho s}(\mathbf{W})$. Then the class is assigned by ML according to:

$$r^\star = arg \max_{r,\rho} \{p_{rs}(\mathbf{W}), p_\rho^e(\mathbf{W})\}.$$

---

[6]For example, http://www.ux.his.no/~tranden/brodatz.html.

[7]The complexity of the Gabor BNT depends on j_0, J, K, so the number of parameters grows linearly with $\log_2(N)$, where N is the size of the sub-image. It may also be sufficient to start with larger j_0 for larger N, so N is critical only for convolution operation and from the statistical viewpoint, to provide a sufficient number of independent training samples.

[8]One can also try to average log-likelihood, instead, or to consider some voting scheme.

Let the number of classification errors for the cross-fold s and texture r to be e_{rs}, then the mean $Err_r = \frac{1}{S} \sum_{s=1}^{S} e_{rs}$ is used to estimate the classification error for texture r.

In addition to CV experiments, the robustness of the proposed models to Gaussian additive noise is checked. A new test image composed of 4 random texture patches with square or triangle layouts (see Figure 2), is contaminated with a small amount of additive Gaussian noise and is presented for classification. Misclassification errors of the texture ensembles for this image are evaluated to assess and compare performances of the different models.

3.1 Texture segmentation results

Four probabilistic models have been simulated in the Gaborian space: the jointly Gaussian probabilistic model (A) and three Gabor BNT models with an increasing complexity: (B) with $\mu = 0$ and with CPTs tied; (C) with $\mu = 0$ and without CPTs tied; (D) without constraints on the CPTs and μ parameters. The Gabor BNTs have been trained by the EM algorithm for a maximum of 200 epochs and are stopped earlier if the log-likelihood increase from one iteration to the next is less than 10^{-4}.

Model's misclassification errors in the CV experiment

Models	D103	D111	D16	D17	D21	D24	D29	D34	D6
A	9	32	13	38	10	35	46	32	14
B	6	31	8	29	8	28	33	25	6
C	7	33	8	25	9	31	39	23	8
D	8	36	11	29	11	38	39	18	9

Table 1: Averaged misclassification errors (in percent) in the cross-validation experiments for different textures (columns) and using different probabilistic models (rows). See text for model's description. Textures are referred to the same as in the Brodatz album.

The mean misclassification error results in percents for CV experiments are presented in Table 1. Our results clearly demonstrate that texture classification by ML using the Gabor BNTs (the lines B-D of Table 1) is superior to using the jointly Gaussian probabilistic model A (the line A of Table 1). It turns out that the most flexible model D is the worst among the considered Gabor BNTs, apart from texture $D34$. This is due to the *curse of dimensionality* problem [27]: there is insufficient data to robustly train a classifier in high dimensional parameter space. This leads to estimators with high variance and large prediction errors. A way to avoid this problem is by imposing appropriate bias constraints or priors [28]. One can easily recognize the model B as a constrained version of the model D. This also demonstrates that the imposed orientation constraints in the Gaborian

space are appropriate. In general, the Gabor model B is the best one for classification.

The robustness results for the models A-B are presented in Table 2. These results are with Gaussian additive noise of $SNR = 20dB$ and $SNR = 40dB$[9]; results without noise ($SNR = \infty$) serve as a baseline for comparison. All misclassification errors are averaged over five runs (each run corresponds to a random noise sampling) and are given in percents. These results clearly show that the Gabor BNT (the model B) is less sensitive to noise and more robust than the model A. In summary, the model B generalizes better than the model A.

Model's robustness to noise

layout	triangle			square		
SNR	∞	$40dB$	$20dB$	∞	$40dB$	$20dB$
Models:						
A:	16.6	18.9	41.4	13.8	16.3	41.2
B:	12.8	14.1	31.3	9.9	11.6	30.8

Table 2: Misclassification errors in percents versus models A, B and versus different levels of Gaussian additive noise. There is no noise when $SNR = \infty$.

4 Texture Detection

In many cases simple generative probabilistic models, such as describing objects by a smoothed intensity or color generated from the normal multivariate distribution is sufficient [29, 30, ignoring the dynamical aspect]. Difficult examples when these models are inappropriate appear due to varying illumination across an image plane or when objects/background are complex textures. In these cases more complex generative models should be applied for image segmentation.

For segmentation by texture in video applications one should construct probabilistic models that are invariant to non-rigid motions, scaling and rotations. The GWT is not invariant to the latter and therefore the stationary, not evolving dynamics, Gabor BNT is only suitable for a constrained class of video images, where the objects are mainly translated and, obviously, for still images.

Segmentation is based on the novelty detection approach [2]. First, the representative part of the object of interest or background is cropped and its probabilistic model is learned. Then the log-likelihood of the image

---

[9]Signal to noise ratio in decibels (dB) is estimated as $SNR = 20\log_{10}\frac{\sqrt{var(x)}}{\sigma}$, where x is a signal with $E[x] \approx 0$ and σ stands for the noise standard deviation.
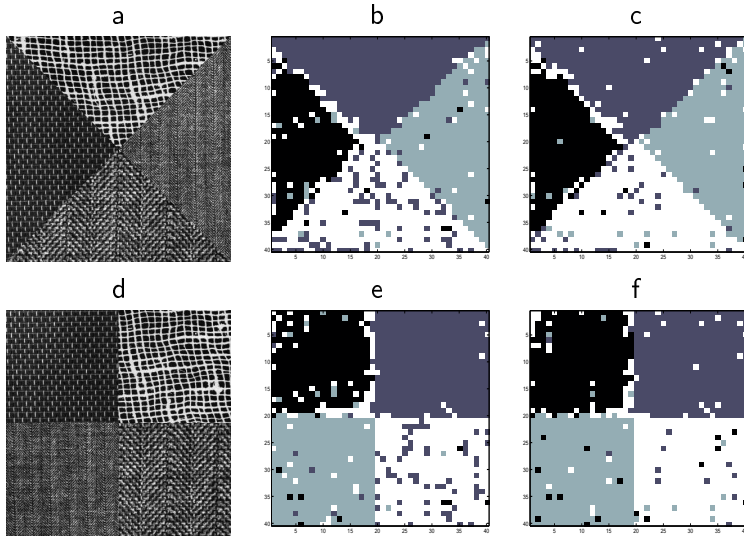
Figure 2: **Robustness experiment**: a. Image composed of textures $D6$, $D103$, $D16$, $D17$ with a triangle layout ($SNR = \infty$); b. Classification results of Figure 2a. with the model A; c. Classification results of Figure 2a. with the model B; d. Image composed of textures $D6$, $D103$, $D16$, $D17$ with a square layout ($SNR = \infty$); e. Classification results of Figure 2d. with the model A; f. Classification results of Figure 2d. with the model B.

pixels is evaluated by the learned model. It is assumed that pixels that do not belong to the learned model get small log-likelihood values under it[10]. Therefore, log-likelihood thresholding can be used for image segmentation. This threshold may be set using cross-validation approach, that is data demanding; instead we use a Gaussian mixture model (GMM) to set a threshold automatically. This means that the log-likelihood of the data is assumed to be moderately well approximated by a mixture of two Gaussian distributions, where the Gaussian component with the larger mean value describes a distribution of the log-likelihood of the data belonging to the learned model. Estimation of the GMM parameters is a standard procedure [31, 32]. The threshold selection procedure for the sea surface pollution detection is schematically illustrated in Figure 3; the problem and data are discussed in the next Section.

In fact, one can go further and attempt to divide novel regions to different levels of novelty using GMM's with more than two mixture components. Intuitively, such segmentation implies that different objects (textures) have

---

[10] Indeed, novelty detection is a semi-heuristical and simplified replacement to the statistical hypothesis testing where the null hypothesis is H_0 : data is generated by the learned probabilistic model and an alternative is H_1 : data is generated by any other model.

different levels of similarity to the learned concept that may be measured by a log-likelihood[11]. This generalization is straightforward in the novelty detection framework. There is no a guarantee, however, that it should work in any possible practical situation.
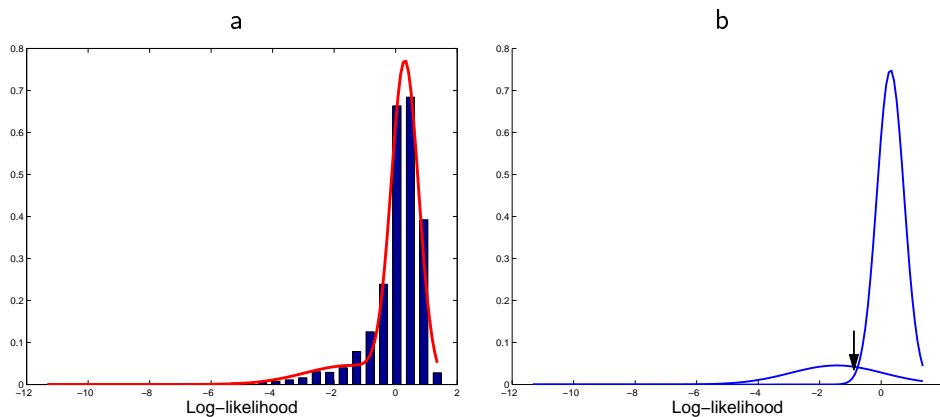

Figure 3: **Threshold selection for novelty detection** : a Empirical distribution (bars) for the log-likelihood of the image pixels and its approximation by the GMM (solid line). b Two Gaussian components properly weighted by their prior probabilities (solid curves) that are estimated by the EM algorithm. The threshold value corresponds to a log-likelihood value in the cross point (marked with the arrow) of the two curves. The learned model is the Gabor BNT for the normal wave region (model C). See Section 4.1 for the problem and data description. The corresponding segmentation result is presented in Figure 4e.

4.1 Sea surface pollution detection

It is well known that water-borne pollutants, natural, such as algal bloom, bacteria and fish oil and leakage from the sea bed, and artificial, such as caused intentionally by ships, generate oily spills (*slicks*) on the water surface. Monitoring and tracking of slick regions is an an important environmental problem. Many government authorities are interested in automatic pollution detection and general assessment of water quality.

As has been demonstrated in [33], the slick regions may be efficiently detected from remotely sensed low-platform mounted visible band camera images. This becomes possible due to different light reflectance of slick and

---

[11]The log-likelihood should be more appropriate than the likelihood in this framework due to the squashing effect of the log-transform. Otherwise, a mixture of the generalized exponential distributions should be more appropriate in the likelihood domain.

surrounding water surfaces and due to different turbulent water motion characteristics of slick and non-slick regions (slicks have a damping effect on the turbulent water motion). As a result, slick regions generally appear brighter in images than normal wave regions (see Figure 4a), as they reflect the sky intensity. This finding has led to a successful *unsupervised segmentation* of sea-surface images based on the Gaussian mixture model (GMM) applied to the tonal (intensity) information [33]. However, this method is very sensitive to illumination varying across the scene [34] and the *main question* is if the slick regions may be detected as textures and not as just bright intensity regions.
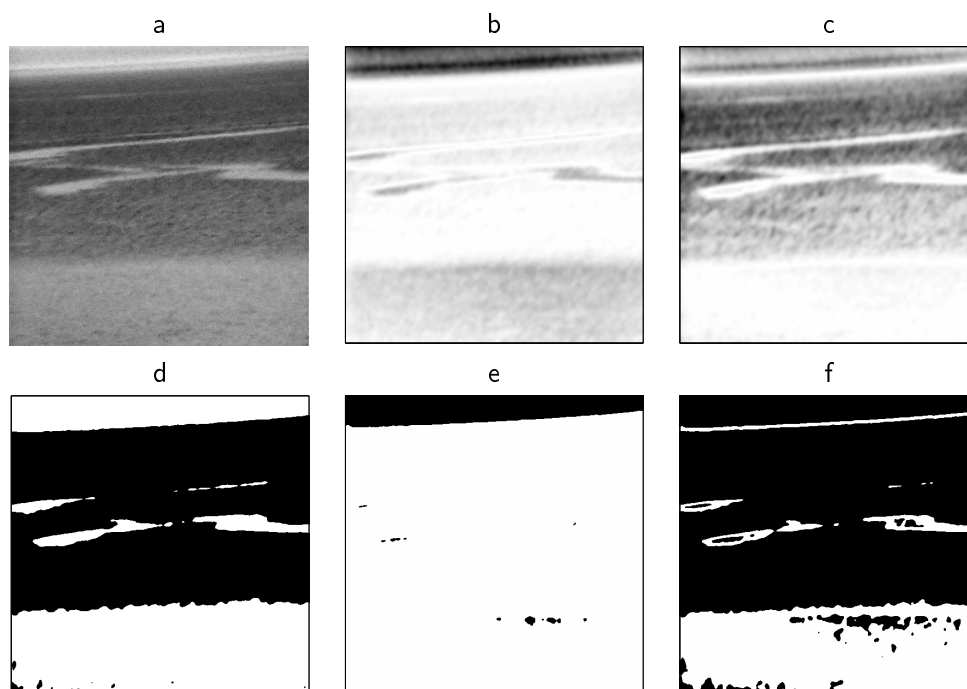

Figure 4: **Gaussian Probabilistic Models:** a. Typical sea surface image; b-c. Log-likelihoods of the image pixels assigned by models (A)-(B); Brighter intensity values correspond to larger values of the log-likelihood; d. Classification by ML using models (A)-(B); e-f. Segmentation based on the novelty detection approach with models (A)-(B), respectively. GMM with two components are applied to set a threshold. White color corresponds to positive examples, i.e. pixels that correspond to sufficiently large log-likelihood values and black color corresponds to novel regions. i.e. to pixels obtaining a small log-likelihood under the normal model.

In order to answer this question, small sub-images corresponding to normal wave and front slick surface regions of the image are cropped and en-

coded by the two types of the probabilistic models. The first type of model are Gaussian models for the low-pass filter coefficients of the wave and slick regions, referred to as models (A) and (B), respectively. These two models are based on the smoothed intensity, they do not carry an information about region textures and are introduced as a baseline for comparison. The other two models (C) and (D) are Gabor BNT models for the wave and slick regions, respectively; they encode GWT coefficients and describe the regions as textures with illumination being partially removed, as low-pass filter coefficients are not modeled by the Gabor BNTs. The Gabor BNT parameters used are: $N = 32$ pixels, octave $j = 2, \ldots, 4$ and 4 orientations.

The image pixel log-likelihoods assigned by the Gaussian models (A)-(B) are presented in Figures 4b-c, respectively. Due to the availability of the probabilistic models for normal wave and slick regions, slick segmentation can be considered as the texture/object classification and detection problems. Image pixel classification by ML based on the models (A)-(B) is presented in Figure 4d. This classification is relatively good, unless the sky region is assigned to the slick region, as sky intensities have large values. Slick detection based on the models (A) and (B) are presented in Figures 4e-f, respectively. Detection based on the model (A) considers the wave regions to be normal (non-polluted) regions and the slick is found as an abnormal (novel) region, that does not belong to a wave region. Novelty detection with the model (A) discriminates the sky and sea surface regions, but does not allow the detection of slicks. Alternatively, detection based on model (B), considers the slicks as being regions of interest and the sky and wave regions emerge as non-interesting regions. In the context of a novelty detection, model (B) appears to be superior to model (A).
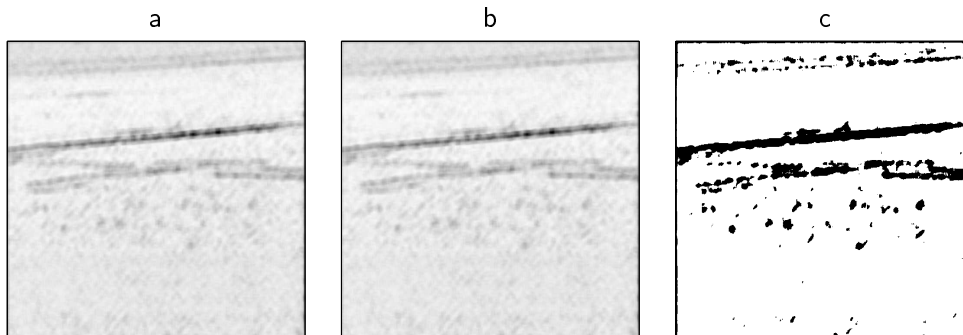

Figure 5: **Gabor BNT models**: a-b Log-likelihoods of the image pixels assigned by the models (C)-(D); Brighter intensity values correspond to larger values of the log-likelihood; c Novelty detection based on the model (C) encoding the normal wave region. A region identified as a new one (negative class) appears in black color.

The log-likelihoods assigned by the Gabor BNTs have been smoothed with a uniform mask of size 11×11 pixels (see Figures 5a-b). This smoothing is equivalent to a product of experts [35, 36] and leads to improved segmentation. The slick detection with the model (C) is presented in Figure 5c, the detection result with the model (D) is similar to the former. Both models identify a narrow slick region as an alien (abnormal) to them. This is due to narrowness of the slick region, i.e. Gabor coefficients are very high around narrow slick edges, so that the narrow slick region pixels get small log-likelihoods under the models (C) and (D). At the same time the models are not able to discriminate between the normal wave and *front* slick regions. *In summary*, the Gaussian models based on intensity features outperform the Gabor BNTs in segmentation. Moreover, the latter have shown to be useful only for edge and the *narrow* slick detection.

To encode Gabor WTC and a low-pass filter coefficient together, the Gaussian models (A)/(B) have been combined with the texture models (C)/(D), respectively. The combined model's probabilities are a direct pixel-wise product of the probabilities assigned to the image pixels by the two models encoding the tonal intensity and texture features. This means that the smoothed intensity and GW coefficients are assumed to be independent. The combined models may be also interpreted as a mixture of two experts. The log-likelihoods of the image pixels assigned by the combined models (Figure 6a.) are very similar to the log-likelihoods assigned by the Gaussian models. Novelty detection by the combined model for the wave region (Figure 6b.) is similar to detection with the Gaussian model (A) (compare with (Figure 4e.), unless edge regions start to appear better as novel regions. The same happens in segmentation with the combined model for the slick region. It may be beneficial to properly weight the log-likelihoods of the mixture models (Gabor BNTs need more weight), but this issue is beyond the scope of our paper.

Results with the extended version of the novelty detection approach with three GMM components for the combined model encoding the wave region is presented in Figure 6c. As can be seen, a new additional cluster includes both slick regions and, unfortunately, the remote wave region; moreover, the image region corresponding to this cluster is identified as more familiar than the sky. We get the same result for Gaussian models (A) and (B). So far additional discrimination of the novelty regions into different novelty (familiarity) levels is quite reasonable.

In summary, the introduced texture models for sea surface slick segmentation allow us to find a narrow slick region, but are not satisfactory in general. This also means that tonal information is the most important clue for slick detection than texture features. An example considered in the next section, in contrast, demonstrates the effectiveness of the proposed probabilistic models.
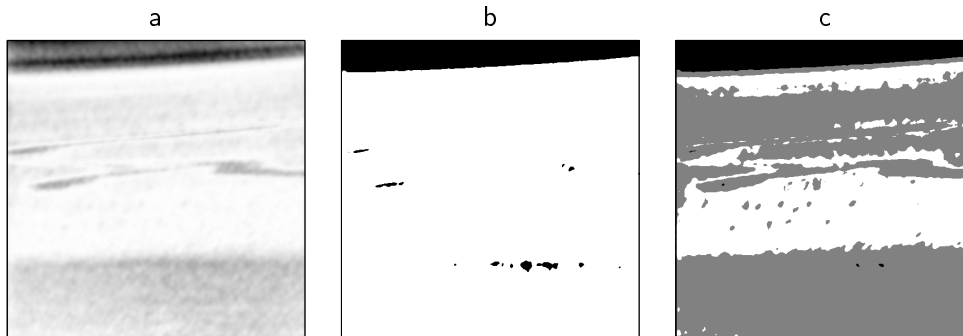
Figure 6: **Combined Gaussian and BNT model for the wave region**:
a Log-likelihood of the image pixels assigned by the joint models (product of
experts (A) and (C)); Brighter intensity values correspond to larger values
of the log-likelihood; b Conventional novelty detection based on the joint
model; c An extended novelty detection based on the joint model with three
Gaussian components; White, grey and black shades stand for regions with
high, intermediate and small log-likelihoods, respectively.

4.2   Texture Detection

This section presents segmentation of a still image with varying illumination
across the scene and textured object/background (see Figure 7a). Part of
an animal body was cropped to build the Gabor BNT. The following Gabor
WT parameters have been used: $N = 32$ pixels, scale $j = 2, \ldots, 4$ and 4
orientations. The Gabor WT coefficients are presented in Figure 7c-d; and
the output coefficients of the corresponding compensating low-pass filter are
presented in Figure 7b. The log-likelihood of the image pixels assigned by
Gaussian probabilistic model for the low-pass filter coefficients is presented
in Figure 7e. The Gaussian probabilistic model has been constructed for a
grass region (the Gaussian model for an animal body region leads to even
worse results as this region does not appear homogeneous). This clearly
indicates that smoothed intensity is not a good feature for segmentation of
this image.

At the same time, an animal body after the Gabor WT appears as a ho-
mogeneous blob. Log-likelihood of the image pixels assigned by the Gabor
BNT learned on the cropped part of the animal body is presented in Fig-
ure 7d. The log-likelihood has been smoothed to satisfy a priori continuity
assumption about it[12]. A result of the log-likelihood smoothing with a uni-
form mask of size 7×7 pixels is presented in Figure 7g and a segmentation
result in Figure 7h.

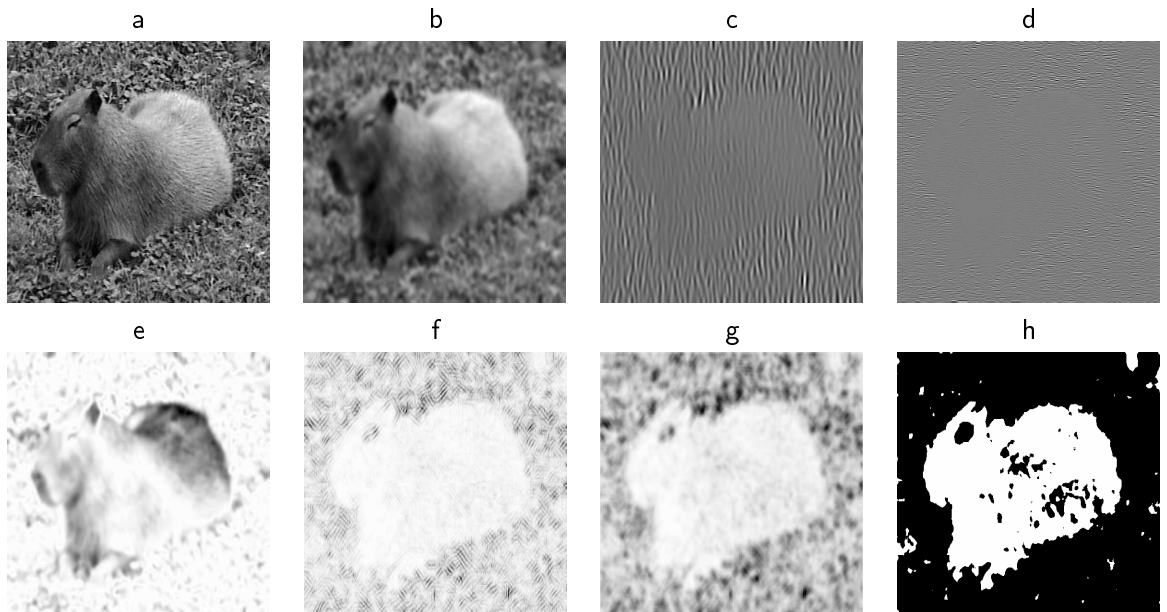[12]This averaging roughly corresponds to a product of experts

Figure 7: a. Original image. b. Output of the GWT compensating low-pass filter. c. The GWT coefficient ($j = 1$ and $k = 1$) as an image. d. The GWT coefficient ($j = 3$ and $k = 3$) as an image. e. Log-likelihood of the image pixels assigned by Gaussian distribution of the smoothed intensity of the grass region. f. Log-likelihood of the image pixels assigned by Gabor BNT learned on the cropped part of the animal body. g. Smoothed log-likelihood. h. Segmentation of the image by thresholding the smoothed log-likelihood.

5   Conclusion and discussion

We have introduced the generative probabilistic oriented wavelet model and have shown that it may be used for texture classification and detection. The introduced model has been compared using cross-validation with the joint Gaussian probabilistic model for several textures from the Brodatz album [1]. Our model is superior to the jointly Gaussian probabilistic model in the Gaborian space, especially when additive noise is added. However, the Gabor BNT training and classification based on it are slower than with the joint Gaussian probabilistic model.

We have studied the feasibility of the introduced generative model for image segmentation in the novelty detection framework [2]. Two examples have been considered: (i) sea surface pollution detection from intensity images and (ii) image segmentation of still images with varying illumination across the scene. The novelty detection framework has been extended by proposing to discriminate the novelty regions into different levels of familiarity based on the GMM operating on the log-likelihood.

19

It may also be interesting to compare this model with the probabilistic model that assumes independent non-Gaussian wavelet coefficient pdfs. Another possibility to encode Gabor WTCs is to use a factorial HMM [37, similar to Fig. 1b]. This BNT better reflects the structure of the Gabor BNT, but is more complex and has only an approximated solution due to loops in the probabilistic graph. Our model may be generalized to address the unsupervised segmentation problem by considering a mixture of the Gabor BNTs. It is interesting to generalize the Gabor BNTs for tracking problems.

Acknowledgments

References

[1] P. Brodatz. *A Photographic Album for Artists and Designers.* Dover, New York, 1966.

[2] C. M. Bishop. Novelty detection and neural network validation. In *IEE Proceedings: Vision, Image and Signal Processing,141, 6,* pages 217–222, 1994.

[3] A. K. Jain. *Fundamentals of Digital Image Processing.* Prentice Hall, London, 1989.

[4] R. C. Gonzalez and P. Wintz. *Digital Image Processing.* Addison-Wesley Publishing Company, 1993.

[5] M. Tuceryan and A. K. Jain. Texture analysis. In C. H Chen, L. F. Pau, and P.S. P. Wang, editors, *The Handbook of Pattern Recogntion and Computer Vision,* pages 207–248. World Scientific Publishing Co., 1998.

[6] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* John Wiley, New York, 1973.

[7] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, New York – Toronto, 1991. third edition.

[8] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 6(6):721–741, 1984.

[9] R. Chellappa and A. Jain. *Markov Random Fields: Theory and Applications.* Academic Press, 1993.

[10] J. K. Romberg, H. Choi, and R. G. Baraniuk. Bayesian wavelet domain image modeling using Hidden Markov trees. In *International Conference on Image Processing - ICIP'99,* pages 1–5, Kobe, Japan, October 1999.

[11] J. S. De Bonet and P. Viola. Texture recognition using a non-parametric multiscale statistical model. In *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition,* 1998.

[12] M. J. Wainwright, E. P. Simonceli, and A. S. Willsky. Random cascades on wavelet trees and their use on analyzing and modeling natural images. *Applied and Computational Harmonic Analysis,* 11(1):89–123, 2001.

[13] S. G. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 1998.

[14] Simoncelli. Statistical models for images: Compression, restoration and synthesis. In *Asilomar Conference on Signals, Systems and Computers*, pages 2–5, Pacific Grove, CA, November 1997.

[15] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, 40:49–71, December 2000.

[16] E. H. Adelson, E. P. Simoncelli, and R. Hingorani. Orthogonal pyramid transforms for image coding. In *SPIE Visual Communications and Image Processing II*, volume 845, pages 50–58, October 1987.

[17] J. G. Daugman. Complete discrete 2D Gabor transforms by neural networks for image analysis and compression. *IEEE Transactions on ASSP*, 36(7):1169–1179, 1988.

[18] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24(12), 1991.

[19] O. Nestares, R. Navarro, J.Portilla, and A. Tabernero. Efficient spatial-domain implementation of a multiscale image representation based on Gabor functions. *Journal of Electronic Imaging, SPIE*, 07(01):166–173, 1998.

[20] D. Marr. *Vision*. Imprint FREEMAN, New York, 1982.

[21] F. V. Jensen. *An Introduction to Bayesian Networks*. Springer Verlag, New York, 1996.

[22] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk. Wavelet-based signal processing using Hidden Markov models. *IEEE Transactions on Signal Processing (Special Issue on Wavelets and Filterbanks)*, April 1998.

[23] M. Jordan, editor. *Learning in Graphical Models*. The MIT Press, Cambridge, Massachusetts, London, England, 1999.

[24] L. R. Rabiner and B. H. Juang. An introduction to hidden markov models. *IEEE ASSP Magazine*, pages 1–16, January 1986.

[25] Kevin Murphy. An introduction to graphical models. Technical report, Intel Research Technical Report., 2001.

[26] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, New York, 1993.

[27] R. E. Bellman. *Adaptive Control Processes*. Princeton University Press, Princeton, NJ, 1961.

[28] S. Geman, E. Bienenstock, and R. Doursat. Neural networks and the bias-variance dilemma. *Neural Computation*, 4:1–58, 1992.

[29] N. Friedman and S. Russel. Image segmentation in video sequences: A probabilistic approach. In *The Thirteenth Conference on Uncertainty in Artificial Intelligence*, pages 1–13. Morgan Kaufmann Publishers, 1997.

[30] Chris Stauffer and W.E.L Grimson. Adaptive background mixture model for real-time tracking. In *IEEE Computer Society Conference om Computer Vision and Patten Recognition, Cat. No. PR00149*, volume 2, page IEEE Comput. Soc. part, 1999.

[31] G. McLachlan and D. Peel. *Finite Mixture Models*. Wiley Seties in Probability and Statistics, New York, 2000.

[32] J. A. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and Hidden Markov models. Technical report, tr-97-021, International Computer Science Institute and Computer Science Division U.C.Berkeley, 1998.

[33] I. Stainvas and D. Lowe. Towards sea surface pollution detection from visible band images. *IEICE Transactions on Electronics a Special Issue on New Technologies in Signal Processing for Electromagnetic-wave Sensing and Imaging*, E84C(12):1848–1856, December 2001.

[34] I. Stainvas and D. Lowe. What is the way to illumination-invariance in visual tasks? Preprint, Aston University, NCRG, June 2002.

[35] T. Heskes. Selecting weighing factors in logarithmic opinion pools. *Advances in Neural Information Processing Systems*, 1998.

[36] G. E. Hinton. Products of experts. In *in Proceedings of the Ninth International Conference on Artificial Neural Networks (ICANN99)*, pages 1–6, Edinburgh, Scotland, 1999.

[37] Z. Ghahramani and M. Jordan. Factorial Hidden Markov models. *Machine Learning*, pages 1–31, 1997.