This is a repository copy of *A Multi-Scale Piecewise-Linear Feature Detector for Spectrogram Tracks*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/67984/

Version: Accepted Version

A Multi-Scale Piecewise-Linear Feature Detector for Spectrogram Tracks

Thomas A. Lampert
Department of Computer Science
University of York, U.K.
tomal@cs.york.ac.uk

Nick E. Pears
Department of Computer Science
University of York, U.K.
nep@cs.york.ac.uk

Simon E. M. O'Keefe
Department of Computer Science
University of York, U.K.
sok@cs.york.ac.uk

Abstract

Reliable feature detection is a prerequisite to higher level decisions regarding image content. In the domain of spectrogram track detection and classification, the detection problem is compounded by low signal-to-noise ratios and high variation in track appearance. Evaluation of standard feature detection methods in the literature is essential to determine their strengths and weaknesses in this domain. With this knowledge, improved detection strategies can be developed. This paper presents a comparison of line detectors and a novel, multi-scale, linear feature detector able to detect tracks of varying gradients. We outline improvements to the multi-scale search strategies which reduce runtime costs. It is shown that the Equal Error Rates of existing methods are high, highlighting the need for research into novel detectors. Results demonstrate that the proposed method offers an improvement in detection rates when compared to other, state of the art, methods whilst keeping false positive rates low. It is also shown that a multi-scale implementation offers an improvement over fixed scale implementations.

1. Introduction

Typically, acoustic data received via passive sonar systems is transformed from the time domain to the frequency domain using the Fast Fourier Transform (FFT). This allows for the construction of a spectrogram image, in which time and frequency are are variables along orthogonal axes and intensity represents the power received at a particular time and frequency. It follows from this that, if a source which emits narrowband energy is present during some consecutive time frames then a track, which is often linear, will be present within the spectrogram. The problem of detecting these tracks is an ongoing area of research with contributions from a variety of backgrounds ranging from statistical modelling [13] and image processing [7, 1, 9] to expert systems [8]. This problem is a critical stage in the detection and classification of sources in passive sonar systems and the analysis of vibration data. Applications are wide ranging and include identifying and tracking marine mammals via their calls [11, 10], identifying ships, torpedoes or submarines via the noise radiated by their mechanics [15, 2], distinguishing underwater events such as ice cracking [4] and earth quakes [6] from different types of source, meteor detection and speech formant tracking [14].

The key step in all of these applications and systems is the detection of the low-level, linear features. Traditional detection methods perform poorly when applied to low SNR images, such as those tested in this paper. Therefore, it is valuable to conduct an evaluation of the standard line detection methods to measure performance, determine weaknesses and strengths which will give insight into the development of novel detection methods for application to this area. We also evaluate the performance of two novel feature detectors, the proposed and a Principal Component Analysis (PCA) supervised learning detector [7].

The problem is compounded not only by the low Signal-to-Noise Ratio (SNR) in spectrogram images but also the variability of the structure of tracks. Structure can vary greatly, including vertical straight tracks, sloped straight tracks, sinusoidal type tracks and relatively random tracks. A good detection strategy should be able to detect all of these.

A variety of standard line detectors have been proposed in image analysis literature, e.g. the Hough transform, Laplacian filter and convolution. There are methods from
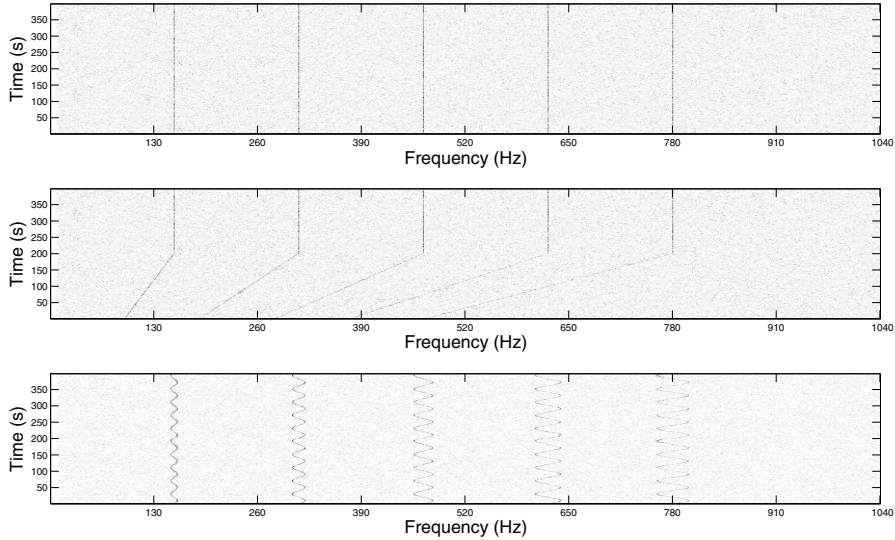
1

Figure 1. Examples of synthetic spectrogram images exhibiting a variety of feature complexities at a SNR of 16 dB.

statistical modelling such as Maximum *A Posteriori* (MAP). Nayar *et al.* [12] describe a more recent parametric detector, proposing that a feature manifold can be constructed using a model-derived training set (in this case a line model) which has been projected into a lower dimensional subspace through PCA. The closest point on this manifold is used to detect a feature's presence within a windowed test image.

This paper is structured as follows: in Section 2 we present the detection methods which have been evaluated with respect to spectrogram images and outline a novel detector. In Section 3, the results of these feature detectors applied to spectrogram images are presented and discussed. Finally, we present our conclusions in Section 4.

2. Method

Three examples of synthetic spectrogram images are presented in Fig. 1, these illustrate the sort of images to which we apply the feature detection methods. The following methods from the literature are applied: the Hough Transform applied to the original grey-scale spectrogram image, the Hough transform applied to a Sobel edge detected image, Laplacian line detection [5], parametric feature detection [12], pixel value thresholding [5], Maximum *A Posteriori* (MAP) detection [3] and convolution of line detection masks [5]. Together with these we also test two novel methods; a detector utilising a bar operator, presented below, and PCA based feature learning which is described in [7]. The parametric feature detection implementation in Matlab was found to be computationally expensive, taking 1.8 hr to perform the detection in one spectrogram. The cause of this was found to be the fine resolution of the parameter variations proposed by the authors to form the manifold. We found it necessary to restrict the manifold to model only

orientation variations which resulted in a large reduction in execution time.

2.1. Bar Operator

Here we describe a simple line detection method which is able to detect linear features at a variety of orientations, translations and scales (width and length) within an image. We propose that this method will also detect linear structure within 2D non uniform grid data, and, can easily be extended to detect structure within 3D point clouds. It can also be easily extended to detect a variety of shapes, curves, or even disjoint regions using different operators.

Initially we outline the detection of an underlying line's angle and subsequently determine its length and width. We define a rotating bar of length l and width w, which is pivoted at one end to a pixel, $\mathbf{g} = (x_g, y_g)$ where $x_g \in \{l, \ldots, N-1\}$ and $y_g \in \{0, \ldots, M - l - 1\}$, in a spectrogram image, $\mathbf{S} = [s_{ij}]_{M \times N}$ (see Fig. 2). The values of the pixels which lie under the bar, $F = \{\mathbf{p} = (j, i) \mid P_l(\mathbf{p}, \theta, l) \wedge P_w(\mathbf{p}, \theta, w)\}$, where

$$P_l(\mathbf{p}, \theta, l) \iff 0 \le [\cos(\theta), \sin(\theta)][\mathbf{p} - \mathbf{g}]^T < l$$
$$P_w(\mathbf{p}, \theta, w) \iff \left|[-\sin(\theta), \cos(\theta)][\mathbf{p} - \mathbf{g}]^T\right| < \frac{w}{2} \quad (1)$$

are summed, such that

$$B(\theta, l, w) = \frac{1}{|F|} \sum_{(j,i) \in F} s_{ij} \quad (2)$$

where θ is the angle of the bar with respect to the x axis.

To reduce the computational load of this calculation, $\mathbf{p} = (j, i)$ in Eq. (1) can be restricted to $j = x_g - (l+1), \ldots, x_g + (l-1)$ and $i = y_g, \ldots, y_g +$
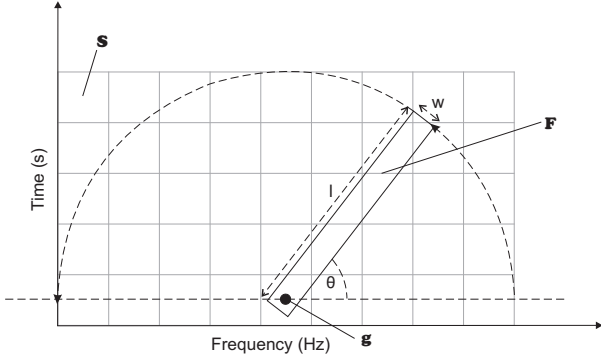
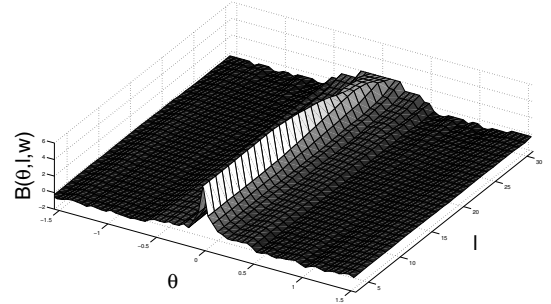Figure 2. The bar operator with width w, length l and angle θ.

Figure 3. The mean response of the bar operator when it is centred upon a vertical line 21 pixels in length (of varying SNRs) and rotated. The bar operator is varied in length between 3 and 31 pixels.

$(l-1)$ (assuming the origin is in the bottom left of the spectrogram) instead of determining $P_w(\mathbf{p}, \theta, l)$ and $P_l(\mathbf{p}, \theta, l)$ for every point in the spectrogram. Also, a set of masks which represent each combination of the parameters θ, l and w can be derived and stored prior to runtime to be convolved with the spectrogram.

To detect the presence and angle of any underlying line the bar is rotated through $180°$ with a fixed width, calculating $B(\theta, l, w)$ at increasing lengths. Normalising the result forms a brightness invariant response, $\bar{B}(\theta, l, w)$ [12], which is also normalised with respect to the background noise, such that

$$\bar{B}(\theta, l, w) = \frac{1}{\sigma(B)} \left[B(\theta, l, w) - \mu(B) \right] \qquad (3)$$

where $\mu(B)$ and $\sigma(B)$ are the mean and standard deviation of $B(\theta, l, w)$.

Statistics regarding the variation of $B(\theta, l, w)$ can be calculated to enable the estimation of an underlying line's angle, $\hat{\theta}$, which passes through the pivoted pixel $\mathbf{g}$. For example, the maximum response can be used, such as

$$\hat{\theta} = \arg\max_\theta \frac{1}{L} \sum_{n=0}^{L} \bar{B}(\theta, l_{\min} + n\Delta l, w) \qquad (4)$$

where $\theta \in \{-\frac{\pi}{2}, -\frac{\pi}{2} + \Delta\theta, -\frac{\pi}{2} + 2\Delta\theta, \ldots, \frac{\pi}{2}\}$, $L = \lfloor (l_{\max} - l_{\min})/\Delta l \rfloor$, Δl and $\Delta\theta$ control the length and angle search resolutions and $w = 1$. A recursive $\arg\max$ can be implemented for the detection of multiple lines. Assuming that noise present in the local neighbourhood of a spectrogram image is random the resulting responses will be relatively low. Conversely, if a line is present, the responses will exhibit a peak in one configuration, as shown in Fig. 3. Comparing the response at the detected angle $\bar{B}(\hat{\theta}, l, w)$ with a threshold t allows the differentiation of these cases, preventing the search for a line's length and width parameters at each pixel if none exists. The threshold will be chosen such that it represents the response obtained when the bar is not fully aligned with a line.

Once the line's angle, $\hat{\theta}$, has been determined and iff $\bar{B}(\hat{\theta}, l, w) \geq t$ we can proceed to analyse $\bar{B}(\hat{\theta}, l, w)$ as l and w are varied, such that $l_{\min} \leq l \leq l_{\max}$ and $w_{\min} \leq w \leq w_{\max}$, to estimate the underlying line's length and width. The response of $\bar{B}$ is dependent on these parameters, as their values increase, and extend past the correct line parameters, it follows that the peak in the response will decrease, illustrated for the length parameter in Fig. 3. An estimate of the length, $\hat{l}$, and width, $\hat{w}$, of the line can therefore be obtained by determining the maximum bar length and width in which the response remains above a threshold value, so that

$$(\hat{l}, \hat{w}) = \max(L_p) \qquad (5)$$

where $\max(L_p)$ is defined as $\max(u, v) = (\max u, \max v)$ where $(u, v) \in L_p$ and

$$L_p = \{(l, w) \mid \bar{B}(\hat{\theta}, l, w) > \frac{3}{4} \max(\bar{B}(\hat{\theta}, l, w))\} \qquad (6)$$

where $l \in \{l_{\min}, l_{\min} + \Delta l, l_{\min} + 2\Delta l, \ldots, l_{\min} + L\Delta l\}$, $w \in \{w_{\min}, w_{\min} + \Delta w, w_{\min} + 2\Delta w, \ldots, w_{\min} + W\Delta w\}$, Δw controls the width search resolution and $W = \lfloor (w_{\max} - w_{\min})/\Delta w \rfloor$. The threshold is taken to be equal to $3/4$ of the maximum response found in $\bar{B}(\hat{\theta}, l, w)$ but could instead be equal to t.

2.1.1 Length and Width Search

The detection of a line's length and width using the linear search outlined above is particularly inefficient and leads to high run-time costs. To reduce this, we propose to replace the uniform search with the more efficient 2D modified binary search algorithm outlined in Alg. 1. Implementing the search in this way reduces the associated search costs from $O(LW)$ to $O(\log L \log W)$, allowing searches to be performed for a large number of line lengths and widths.
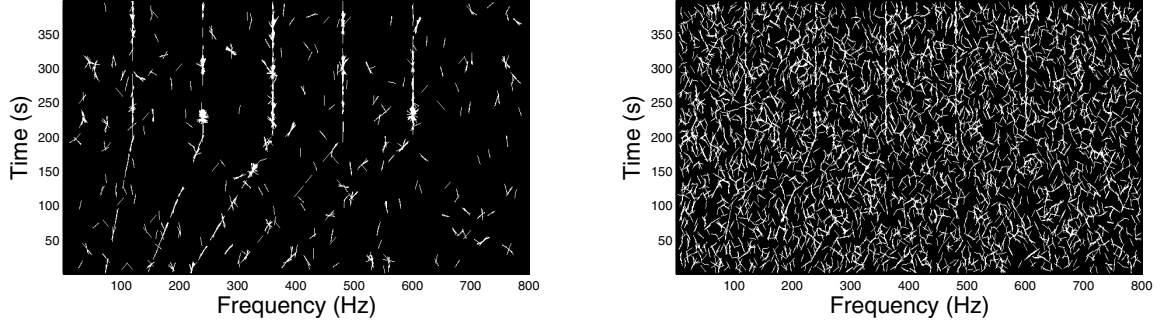
Figure 4. Spectrogram detections (2.18 dB SNR in the frequency domain) using the proposed method (left) and the parametric manifold detection (right).

Algorithm 1 Line parameter binary search.
Input: $l_{\min}$ & $l_{\max}$, the minimum & maximum length to search for, $w_{\min}$ & $w_{\max}$, the minimum & maximum width to search for, Δl and Δw the length and width search resolutions, t, the detection threshold, $\hat{\theta}$, the line's orientation, $\mathbf{S}$, a spectrogram image.
Output: $\hat{l}$ & $\hat{w}$, the length and width of an underlying line.

```
 1:  while l_max − l_min > Δl ∨ w_max − w_min > Δw do
 2:      l ← ⌊ (l_min+l_max)/2 ⌋
 3:      w ← ⌊ (w_min+w_max)/2 ⌋
 4:      if B̄(θ̂, l, w) ≥ t then
 5:          l_min ← l
 6:          w_min ← w
 7:      else {the line's length & width have been exceeded}
 8:          if B̄(θ̂, l, w_min) ≥ t then
 9:              l_min ← l
10:          else {the line's length has been exceeded}
11:              l_max ← l
12:          end if
13:          if B̄(θ̂, l_min, w) ≥ t then
14:              w_min ← w
15:          else {the line's width has been exceeded}
16:              w_max ← w
17:          end if
18:      end if
19:  end while
20:  l̂ ← l_min
21:  ŵ ← w_min
22:  return l̂, ŵ
```

3. Experimental Results

In this section we present a description of the test data and the results obtained during the experiments.

3.1. Data

The methods were tested on a set of 730 spectrograms generated from synthetic signals 200 seconds in length with a sampling rate of 4 kHz (examples of which were presented in Fig. 1). The spectrogram resolution was taken to be 1 sec with 0.5 sec overlap and 1 Hz per FFT bin. These exhibited

SNRs (frequency domain) ranging from -3.5 to 9.5 dB and a variety of track appearances, ranging from constant frequencies (vertical lines), ramp up frequencies (non-vertical lines) (with a gradient range of 1 to 16 Hz/sec at 1 kHz) to sinusoidal (with periods ranging from 10, 15 & 20 seconds and amplitudes ranging from 1–5% of the centre frequency). The test set was scaled to have a maximum value of 255 using the maximum value found within a training set (except when applying the PCA detector as the original spectrogram values were used).

The ground truth data was created semi automatically by thresholding (where possible) high SNR versions of the spectrograms. Spurious detections were then eliminated and gaps filled in manually.

3.2. Nayar Parametric Detection

First we use the feature detector proposed by Nayar *et al.* as a comparison method. This, like the method proposed, is a model based feature detector. The primary difference between the two is that Nayar *et al.* propose to construct a sampled manifold in a feature space derived through PCA. Detection is achieved by calculating the closest point on the manifold to a sample taken from an image and thresholding the distance if necessary. The proposed method performs the detection without the construction of the manifold, instead, the image sample's responses as the model is varied are analysed and the best fit is found from these. This avoids the loss of information that is an effect of dimensionality reduction.

The execution times of the proposed method and that outlined by Nayar *et al.* were measured within one 398×800 pixel spectrogram using Matlab 2008a and a dual-core 2.0 GHz Intel PC. As the comparison method is not multi-scale we fixed $l_{\min} = l_{\max} = 13$ in the bar operator model. Additionally, the parametric manifold was constructed using the same parameter range and resolution as was used with the bar operator. The proposed method performed the detection in 5.5 min whereas the comparison performed the detection in 3.4 min, the resulting detec-
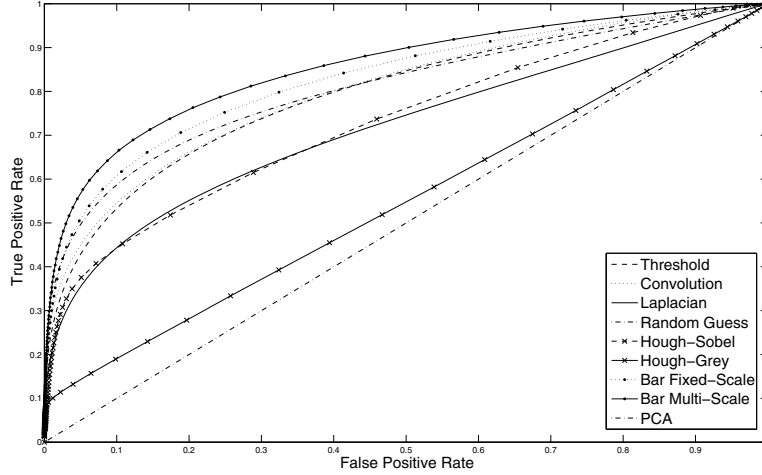
Figure 5. Receiver Operator Curves of the evaluated detection methods.

tions can be seen in Fig. 4. The threshold for each method which resulted in a True Positive Rate (TPR) of 0.7 was found to allow an equivalent False Positive Rate (FPR) to be compared. The comparison method resulted in an FPR of 0.1626 and the proposed method an FPR of 0.0251.

3.3. Results

Here we compare the proposed method to the remaining line detectors using the following parameters. The Laplacian and convolution filter sizes were 3×3 pixels. The threshold parameters for the Laplacian, bar, convolution and pixel thresholding were varied between 0 and 255 in steps of 0.2. Using a window size of 3×21 pixels the PCA threshold ranged from 0 to 1 in increments of 0.001. The bar operator's parameters were set to $w_{\min} = w_{\max} = 1$, $l_{\min} = 6$ and $l_{\max} = 20$. The class probability distributions for the MAP were estimated using a gamma pdf for the signal class and a exponential pdf for the noise class. The PCA method was trained using examples of straight line tracks and noise.

The Receiver Operator Curves (ROC) were generated by varying a threshold parameter which operated on the output of each method - pixel values above the threshold were classified as signal and otherwise noise. The ROC curves for the Hough transforms were calculated by varying the parameter space peak detection threshold. The TPR and FPR were calculated using the number of correctly/incorrectly detected signal and noise pixels.

3.3.1 Existing Methods

The MAP detector highlights the problem of high class distribution overlap and variability; achieving a TPR of 0.0510 and a FPR of less than 0.0002. This rises to a TPR of 0.2829 and a FPR of 0.0162 when the likelihood is evaluated within a 3×3 pixel neighbourhood (as no thresholding

is performed ROCs for these methods are not presented).

It can be seen in Fig. 5 that the threshold and convolution methods achieve almost identical performance over the test set, with the Laplacian and Hough on Sobel line detection strategies achieving considerably less and the Hough on grey scale image performing the worst. We think that the Hough on edge transform outperformed the Hough on grey scale due to the reduction in noise occurring from the application of an edge detection operator. However, both of these performed considerably less well than the other methods due to their limitation of detecting straight lines. The PCA supervised learning method proved more effective than all of these, exceeding the performance of the closest two (thresholding and convolution) indicating that the learning method is capturing the correct type of information. As previously mentioned, the PCA method was trained using vertical, straight track examples only, limiting its sinusoidal and gradient track detection abilities. We think that with extended training, this method could improve further.

3.3.2 Fixed-Scale Bar Operator

Preliminary tests were performed using a fixed length detector. The maximum of the operator's response, $\bar{B}(\hat{\theta}, l, w)$, where $\hat{\theta}$ is defined by Eq. (4) and $l_{\min} = l_{\max} = 21$, was taken as the pixel's value and thresholded to perform the detection. It can be seen in Fig. 5 that the proposed detecor with a fixed-scale bar operator outperforms the methods from the literature.

3.3.3 Multi-Scale Bar Operator

The multi-scale abilities of the proposed method allow it to better fit piecewise linear features and approximate curvilinear features. These properties translate to a ROC curve which has greater separation from existing line detection

methods than the fixed length implementation, and thus it achieves much higher TPRs and lower FPRs. Taking an example TPR of 0.7 the best detectors are, from worst to best; Convolution (FPR: 0.246), PCA (FPR: 0.215), Bar Fixed-Scale (FPR: 0.181) and Bar Multi-Scale (FPR: 0.133). These results show that the combination of intensity and structural information, rather than relying on intensity information alone, increases detector reliability.

4. Conclusions

This paper has presented a performance comparison of line detection methods present in the literature applied to spectrogram track detection. We have also presented and evaluated a novel line detector. The results show an improvement upon results obtained without multi-scale detection and also upon standard line detection methods when applied to this problem. Thresholding is found to be very effective and it is believed that this so because spectrograms with a SNR of 3 dB or more constitute 70% of the test database, circumstances which are ideal for a simple method such as thresholding. However, when lower SNRs are encountered it is believed that thresholding will fall behind more sophisticated methods. Also, thresholding only provides a set of disjoint pixels and therefore a line detection stage is still required. It is noted that the PCA method was trained using examples of straight tracks but was evaluated upon a data set containing a large number of tracks with sinusoidal appearance, reducing its effectiveness but still allowing it to surpass the other existing methods.

When compared to the detection method proposed by Nayer *et al.* the proposed method offers a significant improvement in FPRs when comparable TPRs are achieved in low SNR conditions. This performance improvement is achieved at the expense of a slight increase in execution time. Conducting orientation detection through a multi-scale strategy could possibly reduce this difference.

The evaluation of standard feature detection methods has highlighted the need to develop improved methods for spectrogram track detection. These should be more resilient to low SNR, invariant to non stationary noise and allow for the detection of varying, unknown, feature appearances.

Improving first stage detection methods reduces the computational burden and improves the detection performance of higher level detection/tracking frameworks such as those presented in [7, 13]. A detection method may not outperform others alone, however, it may have desirable properties for the framework in which it is used and therefore, in this case, provide good detection rates.

Acknowledgements

References

[1] J. S. Abel, H. J. Lee, and A. P. Lowell. An image processing approach to frequency tracking. In *Proc. of the IEEE Int. Conference on Acoustics, Speech and Signal Processing*, volume 2, pages 561–564, March 1992.

[2] C.-H. Chen, J.-D. Lee, and M.-C. Lin. Classification of underwater signals using neural networks. *Tamkang J. of Science and Engineering*, 3(1):31–48, 2000.

[3] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.

[4] J. Ghosh, K. Turner, S. Beck, and L. Deuser. Integration of neural classifiers for passive sonar signals. *Control and Dynamic Systems - Advances in Theory and Applications*, 77:301–338, 1996.

[5] R. C. Gonzalez and R. E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[6] B. P. Howell, S. Wood, and S. Koksal. Passive sonar recognition and analysis using hybrid neural networks. In *Proc. of OCEANS '03*, volume 4, pages 1917–1924, September 2003.

[7] T. A. Lampert and S. E. M. O'Keefe. Active contour detection of linear patterns in spectrogram images. In *Proc. of ICPR'08*, pages 1–4, December 2008.

[8] M. Lu, M. Li, and W. Mao. The detection and tracking of weak frequency line based on double-detection algorithm. In *Int. Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications*, pages 1195–1198, August 2007.

[9] J.-C. D. Martino and S. Tabbone. An approach to detect lofar lines. *Pattern Recognition Letters*, 17(1):37–46, January 1996.

[10] D. K. Mellinger, S. L. Nieukirk, H. Matsumoto, S. L. Heimlich, R. P. Dziak, J. Haxel, M. Fowler, C. Meinig, and H. V. Miller. Seasonal occurrence of north atlantic right whale (*Eubalaena glacialis*) vocalizations at two sites on the scotian shelf. *Marine Mammal Science*, 23:856–867, October 2007.

[11] R. P. Morrissey, J. Ward, N. DiMarzio, S. Jarvis, and D. J. Moretti. Passive acoustic detection and localisation of sperm whales (*Physeter Macrocephalus*) in the tongue of the ocean. *Applied Acoustics*, 67:1091–1105, November-December 2006.

[12] S. Nayar, S. Baker, and H. Murase. Parametric feature detection. *Int. J. of Computer Vision*, 27:471–477, 1998.

[13] S. Paris and C. Jauffret. A new tracker for multiple frequency line. In *Proc. of the IEEE Conference for Aerospace*, volume 4, pages 1771–1782. IEEE, March 2001.

[14] Y. Shi and E. Chang. Spectrogram-based formant tracking via particle filters. In *Proc. of IEEE ICASSP*, volume 1, pages I–168–I–171, April 2003.

[15] S. Yang, Z. Li, and X. Wang. Ship recognition via its radiated sound: The fractal based approaches. *J. Acoust. Soc. Am.*, 11(1):172–177, July 2002.