On the Detection of Tracks in Spectrogram Images

Thomas Lampert

On the Detection of Tracks in Spectrogram Images

Thomas A. Lampert*, Simon E. M. O'Keefe*

Department of Computer Science, University of York, Deramore Lane, York, YO10 5GH, UK.

Abstract

This paper proposes an active contour algorithm for spectrogram track detection. It extends upon previously published work in a number of areas, previously published internal and potential energy models are refined and theoretical motivations for these changes are offered. These refinements offer a marked improvement in detection performance, including a notable reduction in the probability of false positive detections. The result is feature extraction at signal-to-noise ratios as low as $-1\,\mathrm{dB}$ in the frequency domain. These theoretical and experimental findings are related to existing solutions to the problem, offering a new insight into their limitations. We show, through complexity analysis, that this is achievable in real-time.

Keywords: Active Contour, Energy Minimisation, Remote Sensing, Spectrogram, Statistical Pattern Recognition, Structural Pattern Recognition

1. Introduction

The problem of detecting tracks in a spectrogram (also known as a LOFARgram, periodogram, sonogram, or spectral waterfall), particularly in underwater environments, has been investigated since the spectrogram's introduction in the mid 1940s by Koenig et al. [21]. Research into the use of automatic detection methods increased with the advent of reliable computational algorithms during the 1980s, 1990s and early 21st century. The research area has attracted contributions from a variety of backgrounds, ranging from statistical modelling [27], image processing [1, 15] and expert systems [25]. The problem can be compounded, not only by a low Signal-to-Noise Ratio (SNR) in a spectrogram, which is the result of weak periodic phenomena embedded within noisy time-series data, but also by the variability of a track's structure with time. This can vary greatly depending upon the nature of the observed phenomenon, but typically the structure arising from signals of interest, can vary from vertical straight tracks (no variation with time) and oblique straight tracks (uniform frequency variation), to undulating and irregular tracks. A good detection strategy should be able to cope with all of these.

*Corresponding author. Tel.: +44 (0)1904 432794; fax: +44 (0)1904 325599.
Email addresses: tomalampert@hotmail.com (Thomas A. Lampert), sok@cs.york.ac.uk (Simon E. M. O'Keefe)

In the broadest sense this problem "arises in any area of science where periodic phenomena are evident" [34]. In practical terms, the problem forms a critical stage in the detection and classification of sources in passive sonar systems, the analysis of speech data and the analysis of vibration data—the outputs of which could be the detection of a hostile torpedo or of an aeroplane engine which is malfunctioning. Applications within these areas are many and include identifying and tracking marine mammals via their calls [26], identifying ships, torpedoes or submarines via the noise radiated by their mechanical movements such as propeller blades and machinery [46], meteor detection, speech formant tracking [39] and so on. Recent advances in torpedo technology has fuelled the need for more robust, reliable and sensitive algorithms to detect ever quieter engines in real time and in short time frames. Also, recent awareness and care for endangered marine wildlife [26] has resulted in increased data collection which requires automated algorithms to detect calls and determine local specie population and numbers.

A spectrogram is a visual representation of the distribution of energy across frequencies and over time, and is formally defined in [25]. The axes of a spectrogram represent discrete time and frequency steps, and intensity represents the amount of power at each time-frequency point. Detecting tracks in spectrogram images is to equivalent to detecting narrowband frequencies in the time domain. It is the concern of this paper to accomplish the former, nevertheless, the differences between the two approaches have been discussed by Wold [44] and reviews of methods which are applied in the time domain have been presented by Kootsookos [22] and Quinn and Hannan [35]. In summary, the transformation of a time domain signal into the frequency domain often allows more efficient analysis to be performed [6]. The transformation also has the effect of quantising a series' broadband noise into the spectrum of frequency bins, and therefore, the SNR of a narrowband feature in the time series is enhanced in the frequency domain [17]. Nevertheless, when constructing a 'conventional' spectrogram image the phase information is lost and, therefore, frequency domain methods should be applied to areas in which the time of measurement commencement is not important.

A recent survey presents a complete review of spectrogram track detection methods [25]. The algorithm presented in this paper captures the salient features of some of these existing methods. By defining an energy minimisation process dependent upon internal energy constraints the active contour can be interpreted as a flexible correlation method. This flexibility avoids the computational complexity that correlation methods [2] require to model equivalent track variations. As such the algorithm incorporates the curvature regularity and temporal continuity features proposed by Di Martino et al. In fact, it is shown that these definitions are insufficient for the variety of track appearances present in this problem and an alternative definition is proposed. Incorporating a gradient potential overcomes the need for initial filtering to direct processing [15]. The model allows for the integration of a priori information regarding harmonic structure to enhance detection rates at low SNRs. Learning optimal filters for the potential energy through statistical analysis of a training set is preferred over employing standard image analysis filters (for example [16, 15]), as these tend

to suppress low SNR tracks and merge closely spaced tracks. Furthermore, these methods, plus those based upon likelihood tests [1, 36], hidden Markov models (HMM) [28, 27] and Dynamic Programming [38], form decisions using information from single pixels; at low SNRs such features become unreliable [24]. Neural Network methods [20] perform low-level feature detection within windows, adding structural information into the decision process. The potential energy proposed in this paper incorporates this to enhance detection probabilities and reduce the probability of false positive detections.

In summary, the contributions presented in this paper can be described as follows:

- A theoretical analysis determines that continuity and curvature measures commonly used in the track detection literature are not suitable and the proposed model is endowed with one that is.
- The existing active contour potential energy is refined to eliminate the threshold parameter and to allow for the removal of tracks from the contour's search space. These improvements simplify the model and reduce the number of false positive detections.
- Discussions regarding the energy minimisation process, rolling windows, and potential energy substitution and augmentation conditions.
- Proof of the algorithm's $O(n^2)$ complexity.
- A thorough evaluation demonstrates the advantages of these developments.
- Limitations of the evaluation metric commonly used in the literature.
- Discussion of this research gives an insight into the weaknesses of existing track detection methods.

The remainder of this paper is organised as follows: Section 2 presents the proposed algorithm; this is evaluated in Section 3, in which a discussion of the results is also presented; and finally the conclusions are drawn in Section 4.

2. Method

This section reviews existing work from the literature and presents a formal definition of the proposed algorithm.

2.1. Related Work

In addition to the overview presented in the Introduction, here are discussed several methods from the literature which have relation to the proposed algorithm.

Di Martino et al. define perceptual track features derived from feature grouping theory [13]. These are: frequential curvature regularity; temporal continuity; high average intensity; and high point density.

Equivalences to the frequential curvature regularity, temporal continuity, and point density features can be found in the proposed algorithm as the curvature and continuity internal energies. Furthermore, the feature described as high pixel intensity is equivalent to the original active contour's potential energy, and both are based upon single pixel values. A recent investigation into low-level feature detection [24] has demonstrated the weakness of such methods when applied to low SNR spectrograms (this finding will be confirmed in Section 3.4.1 of this paper). The same features are also utilised in other work [12], the primary difference between the two being the search strategy employed to group pixels into a track. The first performs an exhaustive search between all pixels [13], evaluating their cost, and selecting those with high values. The second tests the convergence of a self organising map for a track's presence using a similar cost function. The active contour's energy minimisation has a far lower computational burden than the exhaustive search strategy.

Di Martino and Tabbone [14] propose a similar cost function to the active contour that incorporates amplitude and the slope between two pixels (an approximation to the first derivative), but omits curvature. Candidate locations are identified by applying a Gaussian filter and, analogous to the authors' previous work, the cost function groups pixels within these regions into track structures. The algorithm's complexity is reduced by limiting the exhaustive search, however, pixel grouping is dependent upon the accuracy of initial detections. Contrary to this, the active contour's energy minimisation removes the need for limiting the search to identified regions. Furthermore, Gaussian filtering may remove weak tracks from the search space.

Methods utilising the HMM [45, 27] maximise the probability of a track based upon the observation and state transition probabilities. Unlikely track structures are therefore suppressed, however, a solution using multiple parallel track models has also been proposed to expand the range of likely track structures [42]. Its additional complexity is circumvented by the proposed algorithm as its internal energies afford the flexibility to model a wide range of structures. Nevertheless, in terms of the proposed work the HMM can be loosely interpreted as a probabilistic method which learns the form of an internal energy; maximising the probability is tantamount to minimising a cost function. The applied HMMs, however, derive observation probabilities from single pixel values, which, as discussed previously, has been shown to be unreliable [24]. Nonetheless, the proposed potential energy can form a standalone low-level feature detector, which outputs a detection probability. It could therefore be used to increase the reliability of the applied HMM solutions.

Correlation methods such as that presented by Altes [2] are computationally expensive as they form large search spaces. The proposed algorithm can be interpreted as a flexible correlation process, which, because of deformation and efficient energy minimisation removes the computational burden of correlation detectors.

2.2. The Active Contour Algorithm

The Active Contour algorithm is a special form of a Tikhonov stabiliser [41] and can be regarded as regularising the problem [30], it is a deformable energy model proposed by Kass et al. [19] and allows for non-parametric feature detection within an image—ideal for problems in which a priori shape information is not strictly defined. The active contour is driven by an energy minimisation process and is constrained by internal energies, which constrain its shape; these constraints are typically defined in terms of curvature and connectivity. The contour is guided by potential energy, attracting it towards features through local changes in energy gradient, and converges on a minimum of the weighted combination of these energies within the spatial domain of the image.

The original active contour algorithm [19] is a mapping defined on a finite grid (the time-frequency plane) such that $\boldsymbol{v} : \{0, 1, \ldots, k-1\} \rightarrow \{0, 1, \ldots, M-1\} \times \{0, 1, \ldots, N-1\}$ and $\boldsymbol{v}(t) = [m(t), n(t)]$. This forms a deformable contour having k vertices where $m(t)$ and $n(t)$ are functions of the contour vertex's position in the spectrogram $\boldsymbol{S} = [s_{ij}]_{N \times M}$, where $s_{ij} \in \mathbb{R}^+$, such that $m : \{0, 1, \ldots, k-1\} \rightarrow \{0, 1, \ldots, M-1\}$, $n : \{0, 1, \ldots, k-1\} \rightarrow \{0, 1, \ldots, N-1\}$, and $n(t)$ is an increasing injective function. The space of admissible contour deformations $\mathcal{A}$ and the contour's energy $\mathcal{E}$ is defined such that

$$\mathcal{E} : \mathcal{A} \rightarrow \mathbb{R} \tag{1}$$

$$\mathcal{E}(\boldsymbol{v}) = \sum_{t=0}^{k-1} \Big(\alpha |\boldsymbol{v}'(t)|^2 + \beta |\boldsymbol{v}''(t)|^2 + \mathcal{P}(\boldsymbol{v}(t)) \Big) \tag{2}$$

where the terms $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ ($\alpha, \beta \in \mathbb{R}$) control the contour's length and second-order curvature respectively (stabilising functions), and the function $\mathcal{P} : \{0, 1, \ldots, M-1\} \times \{0, 1, \ldots, N-1\} \rightarrow \mathbb{R}$ is the potential induced by the image.

The potential is originally taken to be the image's intensity values, $\mathcal{P}(\boldsymbol{v}(t)) = -\gamma s_{n(t),m(t)}$, or gradient, $\mathcal{P}(\boldsymbol{v}(t)) = -\gamma \left| \nabla s_{n(t),m(t)} \right|^2$, where ∇ is the gradient operator and $0 \leq \gamma \leq 1$ ($\gamma \in \mathbb{R}$) is the image potential's weight.

In an image, the first and second derivatives in Eq. (2), constituting the length and curvature internal energies, are discretely approximated by finite differences [43] such that $\boldsymbol{v}'(t) \approx \|\boldsymbol{v}(t) - \boldsymbol{v}(t-1)\|^2$ and $\boldsymbol{v}''(t) \approx \|\boldsymbol{v}(t-1) - 2\boldsymbol{v}(t) + \boldsymbol{v}(t+1)\|^2$, where $\|\|$ is Euclidean distance. To allow these to be calculated at each end of an open ended contour the vertexes on the opposite side of either end are 'mirrored' through the boundary conditions $\boldsymbol{v}(k) = \boldsymbol{v}(k-2)$ and $\boldsymbol{v}(-1) = \boldsymbol{v}(1)$.

2.3. Track Detection Framework

The limitations of the contour's sensitivity to initialisation and potential energy, in addition to further developments, are addressed in this section. The energy minimisation technique is outlined first, then the internal energy is analysed, and finally the gradient potential and image potential energies are outlined.

It is the concern of this paper to detect vertical curvilinear features. Therefore the described active contour is open ended where the first and last vertices are fixed to the top and bottom of the spectrogram (or window within the spectrogram) such that $\boldsymbol{v}(0) = [m(0), n(0)]$, where $n(0) = \rho$; $\boldsymbol{v}(k-1) = [m(k-1), n(k-1)]$, where $n(k-1) = N - 1 - \rho$; and ρ is the height dimension of the potential energy defined in Section 2.3.4.

2.3.1. Energy Minimisation

The iterative greedy algorithm [43] relaxes constraints on the forms of the energy functions imposed by other minimisation techniques and has a low computational complexity, $O(nm)$. Specifically, the algorithm allows the energy terms to have non-differentiable forms without loss of accuracy [43]. It has also been shown to be quicker than finite differences and LU decomposition [18]. The pseudo-code that describes this process in detail is presented in Algorithm 1. Due to the iterative nature of the process it is possible that it cycles between two states indefinitely and therefore this occurrence should form an additional stopping condition during implementation.

The process updates each contour vertex's position to the minimum within its local neighbourhood. Its neighbourhood is defined such that each vertex remains within the row of the spectrogram in which it is initialised, such that the neighbourhood of $\boldsymbol{v}(t)$ is the set $L_{\boldsymbol{v}(t)} = \{[m(t) - 1, n(t)], [m(t), n(t)], [m(t) + 1, n(t)]\}$ where $\boldsymbol{v}(t) = [m(t), n(t)]$. This restriction preserves the injective properties of the function $n(t)$ and eliminates the need for dynamic resampling [37]. If no movement occurs and the search range has not been exceeded, the contour has converged to a minimum of the energy function $\mathcal{E}$. The contour's position is stored as a detection (see Section 2.4 below), the corresponding positions in the spectrogram are set to zero (see Section 2.3.3 below), and the contour is re-initialised. The search range $f_1, f_2 \in \{1+\phi, 2+\phi, \ldots, M-2-\phi\}$ (the parameter ϕ will be defined as part of the image potential in Section 2.3.4) can be taken as the complete frequency range, i.e. $f_1 = 1 + \phi$ and $f_2 = M - 2 - \phi$ (space is needed at either extreme of the available frequency range for the contour to exceed the search range), or a contiguous subset of the spectrogram corresponding to a frequency range of interest, such that $f_1 < f_2$.

2.3.2. Internal Energy

In this problem there are three ideal track configurations which a detection strategy is required to detect: vertical, a constant frequency; oblique, an increasing or decreasing frequency; and sinusoidal, an undulating frequency (typically a result of the Doppler shift). Therefore, the internal energies of the active contour should be at a minimum in all of these configurations. The internal energies of the previous (and original) active contour model are defined by its first and second derivatives. This is also a common definition presented in track detection literature [13, 12, 14]. Williams and Shah [43] demonstrate that a closed contour under no influence from a potential energy collapses into a point (see also Dilengette's paper on stabilisers [11]). In the open-ended case, it can be easily shown by calculating derivatives that these energies

Algorithm 1 Contour energy minimisation

Input: $\boldsymbol{S}$, spectrogram; f_1 and f_2, search range.

Output: $\hat{\boldsymbol{v}}$, detection positions.

$s \leftarrow 1$
$j \leftarrow 1$
initialise contour $\boldsymbol{v}^s$ at -1
initialise contour $\boldsymbol{v}^{s-1}$ at f_1
while $\forall f_1 \leq m(t) \leq f_2$, where $[m(t), n(t)] = \boldsymbol{v}^s(t)$ **do**
    while $\forall \boldsymbol{v}^s(t) \neq \boldsymbol{v}^{s-1}(t)$ **do**
        for $t = 0, 1, \ldots, k-1$ **do**
            if $f_1 \leq m(t) \leq f_2$, where $[m(t), n(t)] = \boldsymbol{v}^{s-1}(t)$ **then**
                $\boldsymbol{v}^s(t) = [i, j]^T$ where $\underset{(i,j) \in L_{\boldsymbol{v}^{s-1}(t)}}{\arg\min} \mathcal{E}([i, j]^T)$
            else
                $\boldsymbol{v}^s(t) = \boldsymbol{v}^{s-1}(t)$
            end if
        end for
    end while
    if $\exists f_1 \leq m(t) \leq f_2$, where $[m(t), n(t)] = \boldsymbol{v}^s(t)$ **then**
        store $\boldsymbol{v}^s(t)$ in detections such that $\hat{\boldsymbol{v}}^j(t) = \boldsymbol{v}^s(t)$
        $j \leftarrow j + 1$
        for $t = 0, 1, \ldots, k-1$ **do**
            $\boldsymbol{v}^{s+1}(t) = \boldsymbol{v}^s(t)$
            $s_{n(t),m(t)} = 0$, where $[m(t), n(t)] = \boldsymbol{v}^s(t)$
        end for
        $s \leftarrow s + 1$
    end if
end while
return detections $\hat{\boldsymbol{v}}$

bias the contour to take on a vertical straight configuration, thus ignoring the remaining two possible track configurations.

An internal energy, $\mathcal{I} : \{0, 1, \ldots, M-1\} \times \{0, 1, \ldots, N-1\} \rightarrow \mathbb{R}$, based upon local geometric properties of the contour vertices [29] circumvents this problem. In what follows let $t \in \{0, 1, \ldots, k-1\}$. The energy is calculated as the distance from a contour vertex, $\boldsymbol{v}(t)$, to the optimal vertex position, $\boldsymbol{v}'(t)$, such that

$$\mathcal{I}(\boldsymbol{v}(t)) = \|\boldsymbol{v}(t) - \boldsymbol{v}'(t)\|. \tag{3}$$

Next is presented the definition of the optimal vertex position. In their paper, Perrin and Smith present the solution descriptively and not mathematically. In this section a mathematical formulation of the energy is derived; the problem can be formulated as calculating the lengths of the sides of an isosceles triangle, see Fig. 1, and can be solved through simple geometrical properties as follows.

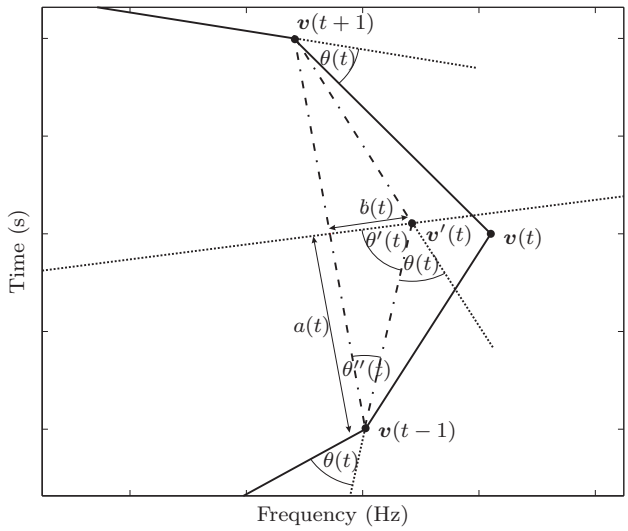


Figure 1: An illustration of the optimal contour vertex position $v(t)'$ for $v(t)$ as defined by the geometric internal energy.

The length of the base of the isosceles triangle is $\|\boldsymbol{v}(t-1) - \boldsymbol{v}(t+1)\|$ and the two equal length sides have lengths $\|\boldsymbol{v}(t-1) - \boldsymbol{v}'(t)\| = \|\boldsymbol{v}(t+1) - \boldsymbol{v}'(t)\|$, where $\boldsymbol{v}'(t)$ is the ideal vertex position. The special case in which the base of the triangle is parallel to the frequency axis will be considered first and the result will be generalised to an arbitrarily rotated triangle. The ideal vertex position for $\boldsymbol{v}(t)$ is at $\boldsymbol{v}(t-1) + [a(t)\ b(t)]$, where $a(t)$ and $b(t)$ are equal to half the length of the isosceles triangle's base and its height, respectively. The component $a(t)$ is therefore calculated such that $a(t) = \frac{1}{2}\|\boldsymbol{v}(t-1) - \boldsymbol{v}(t+1)\|$ and therefore the midpoint between $\boldsymbol{v}(t-1)$ and $\boldsymbol{v}(t+1)$ is simply $\boldsymbol{a}_m(t) = \boldsymbol{v}(t-1) + [a(t)\ 0]$.

The component $b(t)$ is the distance between $\boldsymbol{a}_m(t)$ and the ideal vertex position $\boldsymbol{v}'(t)$ on the perpendicular bisector. This point lies on the perpendicular bisector such that the angle between $\boldsymbol{v}(t-1) - \boldsymbol{v}'(t)$ and $\boldsymbol{v}'(t) - \boldsymbol{v}(t+1)$ is equal to the average angle $\theta(t)$ of the three surrounding contour edges, such that

$$\theta(t) = \frac{1}{3}\cos^{-1}\left(\frac{\boldsymbol{u}(t)\cdot\boldsymbol{u}(t-1)}{\|\boldsymbol{u}(t)\|\,\|\boldsymbol{u}(t-1)\|}\right) \tag{4}$$

where $\boldsymbol{u}(t) = \boldsymbol{v}(t-1) - \boldsymbol{v}(t)$ for any non-zero vector $\boldsymbol{u}(t)$ [31]. Therefore angle $\theta''(t)$ (see Fig. 1) is $\theta''(t) = 180 - 90 - \theta'(t)$ where $\theta'(t) = \frac{180-\theta(t)}{2}$. The length of $b(t)$ is calculated through basic trigonometry, such that $b(t) = a(t)\tan\theta''(t)$ and the ideal vertex position $\boldsymbol{v}'(t)$ is therefore $\boldsymbol{v}(t-1) + [a(t)\ b(t)]$ for the special case that the base of the triangle, having length $2a(t)$, is parallel to the frequency axis. To generalise this, the vector $[a(t)\ b(t)]$ is rotated by θ^* degrees, where θ^* is the angle between the triangle's base and the frequency axis, such that

$$\boldsymbol{v}'(t) = \boldsymbol{v}(t-1) + \begin{bmatrix} a(t) \\ b(t) \end{bmatrix}\begin{bmatrix} \cos(\theta^*) & \sin(\theta^*) \\ -\sin(\theta^*) & \cos(\theta^*) \end{bmatrix} \tag{5}$$

where $\theta^* = \frac{\boldsymbol{u}(t)\cdot[1\ 0]}{\|\boldsymbol{u}(t)\|}$.

This definition replaces the contour's curvature criterion by instead forcing the angles between the contour's edges to be equal. It also replaces the contour's length condition by instead forcing each contour vertex towards a point on the perpendicular bisector of the surrounding vertices.

2.3.3. Gradient Potential Energy

Since introduction the active contour model has been limited by its sensitivity to the initialisation location, an effect of the local nature of the potential energy. In the closed contour case introducing the balloon force overcomes this [9]. Similarly, a gradient potential $\mathcal{W} : \{0, 1, \ldots, M-1\} \times \{0, 1, \ldots, N-1\} \to \mathbb{R}$ is proposed, such that

$$\mathcal{W}(\boldsymbol{v}(t)) = \begin{bmatrix} -c \\ 0 \end{bmatrix} \boldsymbol{v}(t) \tag{6}$$

where $c \in \mathbb{R}$ and $c \geq 0$ (at $c = 0$ the gradient potential is ignored). This forms a uniform gradient in the search space, biasing movement to the direction of increasing frequency with force c. This also prevents the contour from being trapped by isolated edge points, allows it to pass over weak edges [40], enables it to move into concave sections [18] and reduces the contour's sensitivity to its initial configuration.

2.3.4. Image Potential Energy

Chan and Vese [8] state that an edge-detector can be defined by a positive decreasing function $g : \mathbb{R} \to \mathbb{R}$ dependent upon an image's gradient, such that

$$\lim_{z\to\infty} g(z) = 0. \tag{7}$$

The original image potential energy function fulfils this condition but such a simple edge-detector has been shown to be insufficient in this application, where low SNRs are commonly encountered [24]—a problem also encountered by Chalana et al. [7].

A detection mechanism can be defined using machine learning techniques such that it exploits more of the available information [24]. Such a feature detector combines intensity information with spatial information to allow for detection in low SNRs along broken (weak) tracks. In addition to the property defined by Chan and Vese, it is required that the detection mechanism has a low computational burden and correlation methods do not have this property. A potential energy has been defined that has low computational burden, performs favourably in comparison with an exhaustive correlation detector [24], and that fulfils the property defined by Eq. (7) [23]. We modify this potential energy to remove the threshold parameter and to allow detected tracks to be removed from the contour's search space—greatly reducing the potential for false positive detections.

The potential energy is derived from intensity values taken within an $M' \times N'$ pixel window $\boldsymbol{W}_{ij}$, centred on vertex $\boldsymbol{v}(t) = [m(t), n(t)]$ where $i = n(t)$ and $j = m(t)$, such that

$$\boldsymbol{W}_{ij} = \begin{bmatrix} s_{i-\rho,j-\phi} & \cdots & s_{i-\rho,j-1} & s_{i-\rho,j} & s_{i-\rho,j+1} & \cdots & s_{i-\rho,j+\phi} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \iddots & \vdots \\ s_{i-1,j-\phi} & \cdots & s_{i-1,j-1} & s_{i-1,j} & s_{i-1,j+1} & \cdots & s_{i-1,j+\phi} \\ s_{i,j-\phi} & \cdots & s_{i,j-1} & s_{ij} & s_{i,j+1} & \cdots & s_{i,j+\phi} \\ s_{i+1,j-\phi} & \cdots & s_{i+1,j-1} & s_{i+1,j} & s_{i+1,j+1} & \cdots & s_{i+1,j+\phi} \\ \vdots & \iddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{i+\rho,j-\phi} & \cdots & s_{i+\rho,j-1} & s_{i+\rho,j} & s_{i+\rho,j+1} & \cdots & s_{i+\rho,j+\phi} \end{bmatrix}$$

$$\rho = \left\lfloor \frac{N'}{2} \right\rfloor, \qquad \phi = \left\lfloor \frac{M'}{2} \right\rfloor \tag{8}$$

where $M' \in \mathbb{N}$ and $N' \in \mathbb{N}$ are odd numbers defining the size of the window (width and height respectively) such that $\phi \leq j < M - \phi$ and $\rho \leq i < N - \rho$. The intensity values are arranged into a vector $\boldsymbol{V}^{ij}$ having $M'N'$ elements, such that

$$\boldsymbol{V}^{ij} = [s_{i-\rho,j-\phi} \;\; s_{i-\rho+1,j-\phi} \;\; \cdots \;\; s_{i+\rho-1,j+\phi} \;\; s_{i+\rho,j+\phi}]. \tag{9}$$

Finally, PCA is utilised to derive a compact feature vector that represents the window. Testing its membership of a noise cluster gives its similarity to noise. This measure is defined to be maximum when $\boldsymbol{W}_{ij}$ contains a track and minimum otherwise. A multivariate Gaussian distribution $G : \{0, 1, \ldots, M-1\} \times \{0, 1, \ldots, N-1\} \rightarrow [0, 1]$ is used to model the noise cluster and this is fitted to examples of the noise class within a space defined by d principal component basis vectors, such that

$$G(\boldsymbol{v}(t)) \;\; = \;\; \frac{1}{(2\pi)^{\frac{d}{2}} |\hat{\boldsymbol{\Sigma}}^n|^{\frac{1}{2}}} e^{-\frac{1}{2}\hat{\boldsymbol{Q}}^T (\hat{\boldsymbol{\Sigma}}^n)^{-1} \hat{\boldsymbol{Q}}} \tag{10}$$

for $\hat{\boldsymbol{Q}} = \boldsymbol{U}^T \dot{\boldsymbol{V}}_l^{ij}$, where $\hat{\boldsymbol{\Sigma}}^n$ is the standard deviation of the low-dimensional noise cluster, and $\dot{\boldsymbol{V}}_l^{ij}$ is the feature vector after removing the high-dimensional noise cluster's mean.

When deriving the vector $\dot{\boldsymbol{V}}^{ij}$ it is necessary to observe the following condition:

$$\dot{\boldsymbol{V}}_l^{ij} = \begin{cases} 0 & \text{if } \boldsymbol{V}_l^{ij} = 0, \\ \boldsymbol{V}_l^{ij} - \boldsymbol{\mu}_l^n & \text{otherwise,} \end{cases} \tag{11}$$

where $\boldsymbol{V}_l^{ij}$ and $\boldsymbol{\mu}_l^n$ are the lth components of the vectors $\boldsymbol{V}^{ij}$ and $\boldsymbol{\mu}^n$ respectively.

This condition enables the contour to 'ignore' previously detected pixels by setting their values to zero (a physically meaningful condition—if no power is present there is no signal). This is trivial to achieve in the

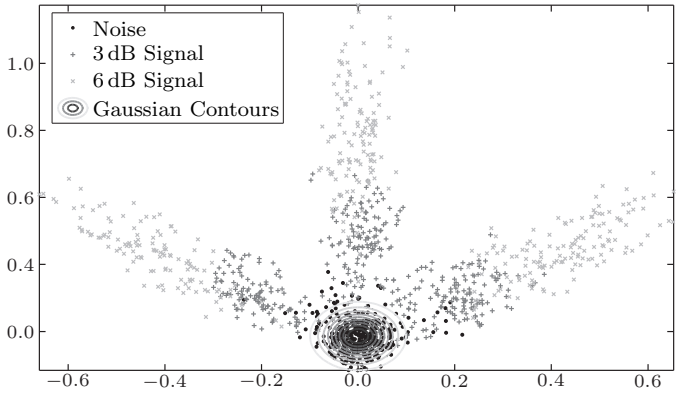


Figure 2: Windowed spectrogram feature vectors projected onto the first two principal components (window size 3×21 pixels, width and height). The noise class is represented by circular points, the two signal classes, having an SNR of 3 and 6 dB, are represented by crosses and the contours represent the Gaussian distribution. Increasing the track's SNR increases its distance from the noise class. The three pronged fan structure results from the track being present in the left, middle or right of the window.

original space but much more involved once the points have been projected into the low-dimensional space. Therefore, the noise cluster is centred on the origin and the condition enforced prior to the projection.

The parameters $\boldsymbol{U}$, $\boldsymbol{\mu}^n$ and $\hat{\boldsymbol{\Sigma}}^n$ are determined as follows and are stored for use during application. The d dimensional subspace's orthogonal bases, $\boldsymbol{U}$, in which the noise cluster, G, will be defined, is derived by applying PCA [4] to a training set $\boldsymbol{X}$. This training set should contain vectors extracted from a spectrogram, some of which contain only noise and some of which contain noise and track, such that $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k]^T$ where $\boldsymbol{x}_k = \boldsymbol{V}^{ij}$.

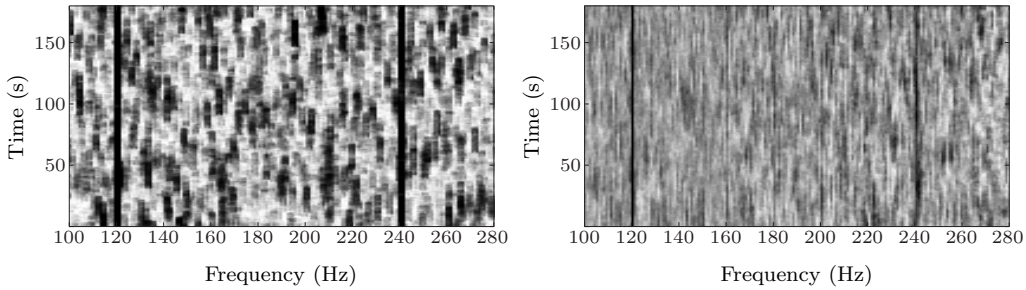
Projecting a training set $\boldsymbol{X}$ onto $d = 2$ basis vectors results in a distribution similar to that presented in Fig. 2; a clear separation of the classes results and noise is clustered using the Gaussian distribution. The level response contours resulting from the Gaussian are depicted in Fig. 2. The value of d is determined during the algorithm's training process.

The noise cluster is fitted to the noise samples in $\boldsymbol{X}$, such that $\boldsymbol{X}^n \subset \boldsymbol{X}$ where $\boldsymbol{X}^n \neq \emptyset$, and their high-dimensional mean is calculated such that

$$\boldsymbol{\mu}^n = \frac{1}{r} \sum_{i=1}^{r} \boldsymbol{x}_i^n \tag{12}$$

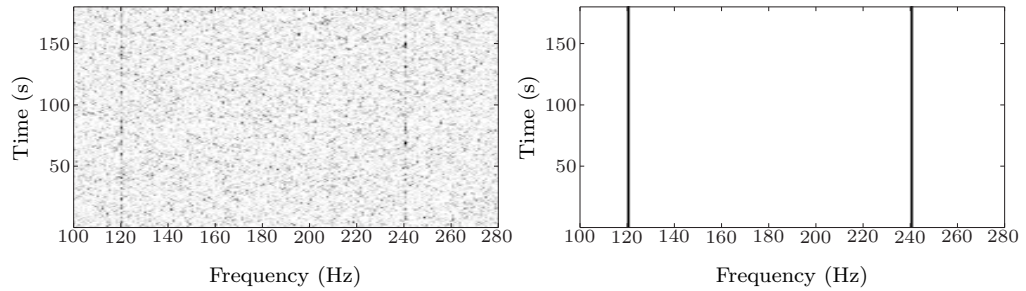
where r is the number of vectors in $\boldsymbol{X}^n$. This mean is removed from the training set to form $\boldsymbol{X}_c^n = \boldsymbol{X}^n - \boldsymbol{\mu}^n$. The standard deviation of the noise-cluster, $\hat{\boldsymbol{\Sigma}}^n$, is then calculated in the low-dimensional space, such that

$$\hat{\boldsymbol{\Sigma}}^n = \frac{1}{r} \sum_{i=1}^{r} (\hat{\boldsymbol{x}}_i^n - \hat{\boldsymbol{\mu}}^n)(\hat{\boldsymbol{x}}_i^n - \hat{\boldsymbol{\mu}}^n)^T \tag{13}$$



(a) Single track, as defined in Eq. (15), $c = 0$.

(b) Multiple tracks, as defined in Eq. (16), $c = 0$ & $\boldsymbol{P}_s = \{1, 2, 3, 4, 5\}$.



(c) Original spectrogram.

(d) Ground truth.

Figure 3: Potential energy 'landscapes' for a 180×180 pixel section of a spectrogram. In the original spectrogram, intensity is proportional to power in voltage-squared per unit bandwidth, that is V^2/Hz. For easier interpretation the values in (a) and (b) are $1 - \mathcal{P}(\boldsymbol{v}(t))$, making the valleys peaks and vice versa. A window size of $M' = 3$, $N' = 21$ (width and height) was used to generate this data.

where

$$\hat{\boldsymbol{\mu}}^n = \frac{1}{r} \sum_{i=1}^{r} \hat{\boldsymbol{x}}_i^n, \tag{14}$$

and $\hat{\boldsymbol{x}}_i^n \in \hat{\boldsymbol{X}}^n = \boldsymbol{U}^T \boldsymbol{X}^n$ (for $i = 1 \dots r$).

Single Track Detection. The output of Eq. (10) is combined with the gradient potential to replace $\mathcal{P}$ in Eq. (2), such that

$$\mathcal{P}(\boldsymbol{v}(t)) = \mathcal{W}(\boldsymbol{v}(t)) + \gamma G(\boldsymbol{v}(t)). \tag{15}$$

where $\gamma \in \mathbb{R}$ and $0 \leq \gamma \leq 1$. Furthermore, the output of Eq. (10) can be utilised as a separate transformation and applied to the spectrogram prior to the active contour (it is therefore compatible with any minimisation technique that is applicable to the original image potential).

The image potential 'landscape' resulting from Eq. (15) is similar to that presented in Fig. 3a, demonstrating that the combination of spatial and intensity information produces few spurious detections and a large gradient change at track locations. Nevertheless, if each simultaneous track is a component originating from a common source, i.e. part of a signature pattern, the energy term results in multiple detections, one

for each component, not a single detection of the whole pattern. In addition to this some false positive detections have similar magnitudes as true positives which are, in short time frames, are hard to separate with a simple threshold.

Multiple Track Detection. To overcome these issues it is necessary to augment low-level information with harmonic information. Integrating harmonic locations in the spectrogram causes a single, more distinguished, response and suppresses false positive detections. This is integrated into the potential energy by extending the notion of harmonic positions to those defined by the characteristics of the target to be detected. As such, the harmonic transformation is a special case of this in which the relationships are defined by integer multiples.

Simultaneous tracks originating from a common source can have some underlying linear relationship, for example, periodic signals are made up of harmonic frequencies and produce tracks at harmonic locations. This a priori knowledge can be represented by a pattern set $\boldsymbol{P}_s = \{a_1, \ldots, a_h\}$, where $a_i \in \mathbb{R}^+$, is a multiple of the fundamental frequency, and can be integrated into the potential energy function, Eq. (15), such that

$$\mathcal{P}(\boldsymbol{v}(t)) = \mathcal{W}(\boldsymbol{v}(t)) + \frac{\gamma}{h}\left[\sum_{i=1}^{h} G\left(\begin{bmatrix} a_i & 0 \\ 0 & 1 \end{bmatrix}\boldsymbol{v}(t)\right)\right] \tag{16}$$

where $a_1 = 1$ (the fundamental frequency), the term $h \geq 1$ is the number of relative frequencies in $\boldsymbol{P}_s$, $\boldsymbol{v}(t) = [m(t), n(t)]$, and $\phi \geq a_i m(t) < M - \phi$ for $i = 1 \ldots h$. The active contour thus forms a pattern-based search.

This modified feature space is presented in Fig. 3b. The averaging of detections vastly reduces the unwanted effects of the energy term defined in Eq. (15). In particular, the response is now located at a single, localised, position corresponding to the fundamental frequency and this is easily distinguished from the harmonic response. Gaps in the track have been interpolated with information from higher harmonics and false positives are weaker due to the random, uncorrelated, nature of noise.

Some Notes on Noise. The following considerations regarding noise modelling could be beneficial to future research endeavours.

A single Gaussian distribution is used to cluster noise in the proposed algorithm. Excursions tending towards the edge of this cluster are more likely to be interpreted as signal, which improves generalisation, particularly when the feature's appearance varies significantly against a background of uniform noise. Within this framework noise with different characteristics may be modelled; in a similar vein to the generalisation made in Eq. (7), any noise model that can be defined as a positive increasing function and is dependent upon the noise in a sample can be adopted—a criterion that all probabilistic models fulfil.

A well defined track class, i.e. where it has little variability, allows for the noise model to be augmented with information derived from the track's characteristics. In this circumstance the term G in Eq. (16) should

be supplemented such that

$$G(z) = 1 - s(z), \tag{17}$$

where $s : \mathbb{R} \to [0, 1]$ is a positive and increasing function of membership to the track class.

Some problems may exhibit temporal variability of the noise distribution—a characteristic that is currently not explicitly accounted for. Nevertheless, an extension can be incorporated to account for this, although it is not fully explored in this paper. Two solutions will be briefly presented. The first solution is to subtract the mean from the window vectors prior to processing, similar to the technique of achieving lighting invariance in photometric image analysis. The second exploits the sparseness of frequency tracks in a spectrogram; in most cases a source will not be present, however, in the case where there exists a source that emits ten narrowband frequency components and the spectrogram is the result of a 1000 bin FFT process, the tracks represent only 1% of the data. They therefore have a very small skew on the estimate of the data's mean and could be accounted for through bias estimation. As such, the mean of the noise cluster could be updated to the (unbiased) mean value of the current spectrogram frame.

A Note on the Vertices' Neighbourhood. As the potential energy is formulated to make use of windows of pixels, the windows' positions with respect to the vertex neighbourhood should be considered. The original potential energy effectively utilises windows of size 1×1. Extending this to a larger size results in windows centred upon each point in the vertex neighbourhood, $L_{\boldsymbol{v}(t)} = \{[m(t) - 1, n(t)], [m(t), n(t)], [m(t) + 1, n(t)]\}$ where $\boldsymbol{v}(t) = [m(t), n(t)]$, and these windows, having a width of three pixels ($M' = 3$, Eq. (8)), will extend from $m(t) - 2$ to $m(t)$, $m(t) - 1$ to $m(t) + 1$, and $m(t)$ to $m(t) + 2$ respectively. This configuration results in information being derived from overlapping positions in the spectrogram. To overcome this, j in Eq. (8) should be taken to be $j = m(t) - 1 - \phi$ if the neighbourhood position is $[m(t) - 1, n(t)]$, and $j = m(t)+1+\phi$ if the neighbourhood position is $[m(t)+1, n(t)]$ (and remains $j = m(n)$ for the neighbourhood position $[m(t), n(t)]$). The resulting configuration means that the contour's evolution is driven by the correct information, i.e. the forward motion is driven by information derived strictly from in front of the contour.

2.4. Rolling Window

It is not always the case that a spectrogram is a stationary image, it can also be constructed in real-time. At each time step the oldest row of the spectrogram is removed and a new prepended, forming a "waterfall display". Track detection can be repeated within this window as the data is updated and leads to consideration of how often the detection is performed and how detections (the positions in each rolling window) are interpreted. Several interpretations are possible, in this paper the following schema is utilised.

Detection is performed as each row is prepended, and therefore the detection process integrates past (and perhaps future) information to enhance the detection at each time point. In this way each row supports k detections as it flows down a rolling window k pixels in height and the contour has k vertices. A set

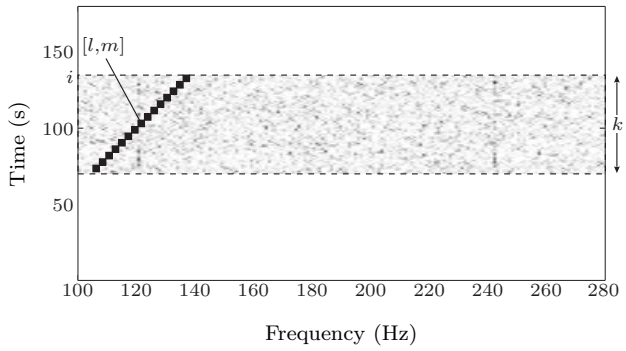


Figure 4: An illustration of a rolling window where $[l, m]$ is the mean position of the detection, k is the height of the window, and i is the position of the window in the spectrogram. In this example, and referring to Eq. (18), $D_t = \{[l, m]^T\}$.

of detections, D_t, comprises the mean positions (along the frequency axis) of the contour vertices in each detection within the rolling window, such that

$$D_t = \left\{ [l,m]^T \middle| [l,m]^T = \left[i - \left\lfloor \frac{k}{2} \right\rfloor, \frac{1}{k} \sum_{t=0}^{k-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \hat{\boldsymbol{v}}_i^j(t) \right]^T \right\} \tag{18}$$

where $i = k, k+1, \ldots, N-1$ is each row of the spectrogram and $\hat{\boldsymbol{v}}_i^j(0)$ is the location in the first row of the jth detection within the rolling window that has row i of the spectrogram as its first row.

2.5. Complexity Analysis

An analysis of complexity is presented using big O notation, where n is the number of elementary arithmetic operations. It has been shown that the greedy energy minimisation algorithm has complexity $\mathrm{O}(mn)$ [43].

The modifications to the potential energy all have a linear complexity and therefore the complexity remains $\mathrm{O}(n^2)$ due to the matrix multiplication between $\boldsymbol{Q}^T$ and $\boldsymbol{\Sigma}^{-1}$ [23]. What remains to be shown is that the geometric internal energy formulation does not introduce greater complexity. It can be seen that the only operations of the geometric energy that are not linear are: cos, *sin*, $\cos^{-1}$, $||\,||$ and the matrix multiplication in $\boldsymbol{v}'(t)$. Analysing these in turn, the trigonometric functions, cos, sin, and $\cos^{-1}$, can all be computed in $O(M(m)\log(m))$, where m is the number of digits precision and $M(m)$ is the number of single precision operations required to multiply m-bit integers [5], assumed to be constant in this analysis. The square and square root involved in the distance function $||\,||$ can all computed in $\mathrm{O}(M(m))$ using Newton's method [5]. The matrix multiplication involved in calculating $\boldsymbol{v}'(t)$ can be achieved using schoolbook multiplication, in $\mathrm{O}(mnp)$ [10], where m, n and p are matrix dimensions, in this case $m = 1$, $n = 2$ and $p = 2 = n$ resulting in a complexity of $\mathrm{O}(n^2)$ where $n = 2$ as these matrices are of fixed size. This

operation has the highest complexity within the algorithm, which is equal to that of the original formulation, it is therefore concluded that the geometric internal energy introduces no additional complexity.

3. Experimental Results

To promote the dissemination of results and the comparison of methods, an implementation of this algorithm and the data set used are available on-line (`http://stdetect.googlecode.com/` and `http://code.google.com/p/stdetect/wiki/DataSets/`).

3.1. Data

The data set used within this paper consists of 4142 spectrogram images. They are formed by generating synthetic acoustic signals and transforming these into the frequency domain. The acoustic time series contain a fundamental frequency of $f_0 = 120\,\text{Hz}$ (at constant speed), a harmonic pattern set $\boldsymbol{P}_s = \{1, 2, 3, 4, 5\}$, and have a sampling rate of $f_s = 4\,\text{kHz}$. Spectrograms are generated using a time resolution of one second with a half second overlap, and a frequency resolution of 1 Hz per STFT bin. The three variations of track appearance that are commonly seen in this problem are: sinusoidal, representing a Doppler shifted signal; vertical, representing a constant engine speed; and oblique, representing an accelerating engine (or a Doppler shift). A number of noise-only spectrograms were also included in the data set. A description of the parameter variations used for these three signal types is outlined in Table 1. For each parameter combination, ten spectrograms are generated to form a test set, and two spectrograms to form a training set. The parameters described in Table 1 are defined as:

Period the time in seconds between two peaks of a sinusoidal track;

Centre Frequency Variation the amplitude of a sinusoidal track relative to its frequency location, expressed as a percentage of the track's frequency;

SNR the frequency domain SNR, described by $\text{SNR} = 10\log_{10}\left(\frac{\bar{P}_t}{\bar{P}_b}\right)$ where $\bar{P}_t = \frac{1}{|P_t|}\sum_{(i,j)\in P_t} s_{ij}$, $\bar{P}_b = \frac{1}{|P_b|}\sum_{(i,j)\in P_b} s_{ij}$ and where $P_t = \{(i,j)|s_{ij}$ belongs to a track$\}$ such that $P_t \neq \emptyset$ and $P_b = \{(i,j)|(i,j) \notin P_t\}$ is the set of points which represent noise such that $P_b \neq \emptyset$;

Track Gradient the amount of change in the track's frequency relative to time.

Ground truth data is created by generating high SNR (approximately 1000 dB) spectrograms, and thresholding these to obtain binary bitmaps.

Track Type	Parameter	Values
Vertical	Signal Duration (seconds)	100
	SNR (dB)	−1–7
Oblique	Track Gradient (Hz/s)	1, 2, 4, 8 & 16
	Signal Duration (seconds)	100
	SNR (dB)	−1–7
Sinusoidal	Period (seconds)	10, 15 & 20
	Centre Frequency Variation (%)	1, 2, 3, 4 & 5
	Signal Duration (seconds)	200
	SNR (dB)	−2–6

Table 1: Parameter values spanning the synthetic data set.

3.2. Evaluation Metrics

Two evaluation measures are adopted, the first being the Line Location Accuracy (LLA) proposed by Pratt [32], and used by Di Martino and Tabbone [15] for this application. A set of all detected pixel locations D_t is constructed, Eq. (18), and there also exists ground truth data in the form of a set of actual pixel locations $P_t = \{(i,j)|s_{ij} \text{ belongs to a track}\}$. The figure of merit is formulated such that

$$F = \frac{1}{\max(|P_t|,|D_t|)} \sum_{(l,m)\in D_t} \frac{1}{1+\lambda \min\limits_{(i,j)\in P_t} (||[l,m]-[i,j]||^2)} \tag{19}$$

where $|P_t|$ and $|D_t|$ are the cardinalities of the actual and detected track pixel set respectively, $||[l,m]-[i,j]||$ is the Euclidean distance between the detected track pixel $[l,m]$ and the actual track pixel $[i,j]$ and $\lambda \in \mathbb{R}^+$ where, throughout this paper, $\lambda = 1$. It can be observed that the LLA measure aggregates a number of commonly measured factors into one figure of merit. To gain further understanding of the balance between these aggregated factors the second evaluation measure will be the mean number of false positive detections.

It is not possible to directly compare the presented LLAs to those of Di Martino and Tabbone [15] as the value of λ used to derive their results is not known.

3.3. Parameter Selection

Preliminary experiments indicated that calculating the PCA vectors using low SNR data samples maximises the signal detection rate. The lower dimensional subspace in which the noise is clustered using Eq. (10) is therefore derived using PCA and 1000 feature vectors containing noise and 1000 feature vectors containing track and noise, each extracted from within a 3×21 pixel window (width and height) from spectrograms having a mean SNR of $-0.5\,\text{dB}$. The PCA eigenvalues, which are presented in Fig. 5, show that the first three account for the largest portion of the data's variance and, therefore, a PCA dimension of three is used throughout these experiments.

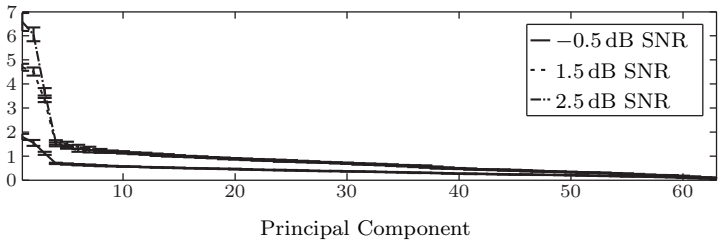
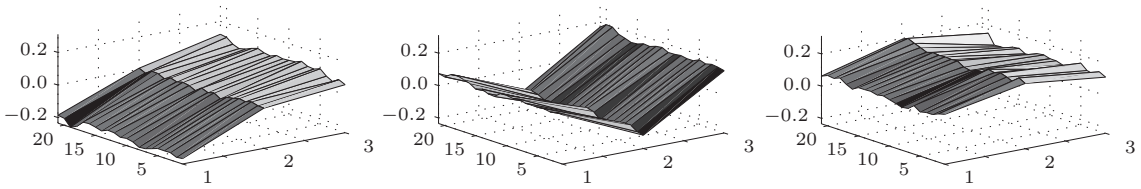


Figure 5: The eigenvalues associated with the principal components derived by averaging over 10 random training sets, each containing 1000 examples of each class. The top line represents the eigenvalues for 2.5 dB SNR examples, the middle 1.5 dB SNR and the bottom −0.5 dB SNR and error bars of 2 standard deviations.



(a) First Principal Component. (b) Second Principal Component. (c) Third Principal Component.

Figure 6: The first three principal component vectors viewed as 3×21 point surface plots.

Reconstructions of the principal component vectors, Fig. 6, confirm PCA's ability to capture salient information in the data. The first is similar to the Prewitt edge detector [33]; the second, a second partial differential edge detector, similar to $s''_{ij} = s_{i-1,j} - 2s_{ij} + s_{i+1,j}$; and the third the inverse of that.

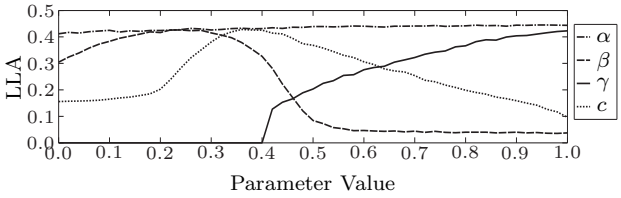
The search was optimised by initialising the contour within 10% of the expected frequency position. Throughout the experiments the active contour's length is set to $k = 20$ and SNRs are rounded to the nearest 0.5 dB.

3.4. Experimentation

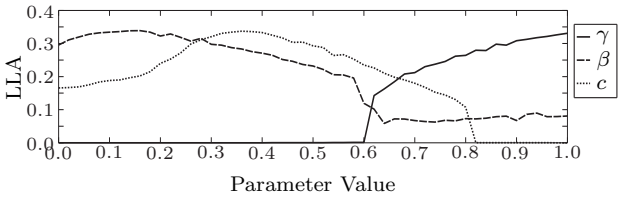
The algorithm's sensitivity to varying weighting values is analysed, a suitable combination of weights for each is selected, and the evaluation is presented.

3.4.1. Parameter Sensitivity

Ballerini explains that "large values for the continuity [length] and curvature weights will discourage convergence to a 'busy' contour" and that "small weights may allow the contour to be trapped into false edges or leak out through gaps in the boundary" [3]. Here suitable parameter values are searched for using the training set by varying each parameter in turn throughout its range of values, whilst the remaining parameters are fixed at values that have lead to good convergences during preliminary experimentation. In



(a) Original internal energy algorithm variant. The mean standard deviations for each function are $\alpha = 0.0042$, $\beta = 0.0052$, $\gamma = 0.0044$ and $c = 0.0057$. Whilst varying each of the parameters the remainder took the following values: $\alpha = 0.10$, $\beta = 0.20$, $\gamma = 1.00$ and $c = 0.41$.



(b) Geometric internal energy algorithm variant. The mean standard deviations for each function are $\beta = 0.0062$, $\gamma = 0.0028$ and $c = 0.0047$. Whilst varying each of the parameters the remainder took the following values: $\beta = 0.20$, $\gamma = 1.00$ and $c = 0.41$.

Figure 7: The mean LLA of training set detection as functions of each variant of the algorithm's parameter values. The results were obtained from five repetitions of the experiment using the potential energy proposed in this paper.

this search, therefore, the interactions between the energies that these parameter values control are ignored and the result is likely to be sub-optimal, however, it greatly simplifies the optimisation process.

Two variants of the algorithm will be evaluated, one will use the original internal energy and the other the geometric internal energy discussed in Section 2.3.2. Presenting the LLA as a function of each parameter's value gives insight into the algorithm's sensitivity to parameter values, Fig. 7.

As the influence of the potential energy is increased the active contour gains more information from the spectrogram. This is reflected in both of the observed functions; as γ increases the LLAs also increase.

The gradient potential parameter c enables the active contour to locate features that lie outside its local gradient and to pass over false positive detections. It can be seen that as c increases, the LLAs also increase, however, if it's value is too great (above 0.36) the contour begins to be forced over true positives and the LLA decreases. In both variants of the algorithm, the functions of c have similar form, both peaking at the same value, indicating that the gradient potential balances the effects of the potential energy and is, in the most part, independent of the contour's internal energy.

The internal energy parameters control the contour's ability to deform and to model the track's structure. When the original internal energy variant is considered, it can be observed that the value of α (which controls the length and parameterisation of the contour) has very little effect on the detection rate. The parameter β (which controls the curvature of the contour) has a far greater effect: at low values the contour has sufficient freedom to model track variations and evolve, however, when the influence is too great (above a value of approximately 0.4) it becomes the dominant energy and the contour is not able to evolve. A similar behaviour is observed when using the geometric internal energy.

If these results are compared to those obtained using the active contour's original potential energy, Fig. 8,

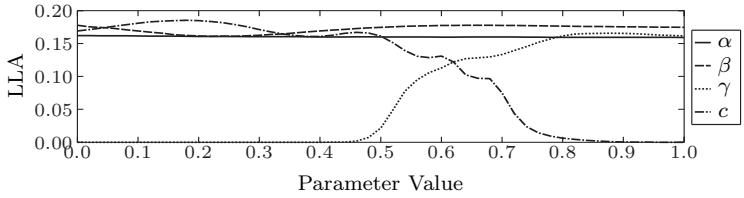


Figure 8: The mean LLA of the training set detections as functions of each variant of the algorithm's parameter values. The results were obtained using the original potential energy (the pixel's intensity). Whilst varying each of the parameters the remainder took the following values: $\alpha = 0.10$, $\beta = 0.20$, $\gamma = 1.00$ and $c = 0.41$.

we see that the original potential energy cannot match the performance achieved by the proposed potential energy. Moreover, the performance actually decreases as the influence of the potential energy is increased, indicating that it is not suitable for this problem.

For the remainder of the experiments the parameter values are set to $\alpha = 0.96$, $\beta = 0.22$, $\gamma = 1.00$ and $c = 0.36$ for the original internal energy variant; and for the geometric internal energy variant: $\beta = 0.16$, $\gamma = 1.00$ and $c = 0.36$. The LLAs presented are the average of ten repetitions of the experiment.

3.4.2. Performance

The LLA of each variant is measured as a function of the SNR of each track configuration allowing for the performance to be analysed as the track's SNR degrades. The mean number of false positive detections is calculated as the mean over all SNRs; false positive detection is dependent upon noise, and not the strength of a track, and it is therefore not a function of SNR. The mean false positive rates are accompanied by significance values, which have been calculated using the rank-sum test and represent the significance that the results from each algorithm variant derive from identical distributions, to 99% confidence (low values indicate that the difference between the results is statistically significant).

The first of these comparisons, with regards to the vertical track structure, is presented in Fig. 9. At SNRs of 0 dB and above the geometric variant outperforms the original by a mean LLA measure of 0.0260. Below this point the performance of both variants degrade, however, the original internal energy is marginally more resilient and has an LLA measure that is, on average, 0.0298 higher than the geometric internal energy variant in the SNR range -1 to -0.5 dB. This is possibly an effect of the shape bias that is exhibited by the original internal energy, as discussed in Section 2.3.2. It is most likely, however, to be a result of the increased number of false positive detections that the original algorithm exhibits. If these are close to the true location the LLA score will artificially increase. To investigate this a comparison between two example detections is presented in Fig. 10, it is clear that the geometric internal energy produces a more favourable result, however, because of the additional false positive detections the original internal energy results in a

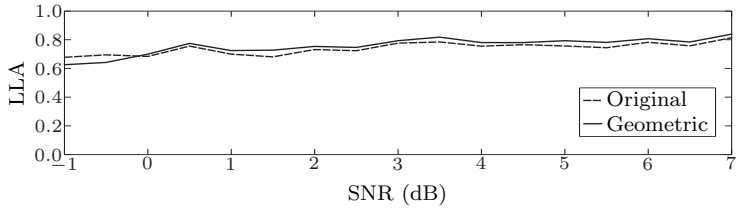


Figure 9: The mean line location accuracies of vertical track detections as functions of the spectrogram's SNR. The mean standard deviations for the Original detections is 0.0060 and the Geometric detections 0.0064.

Original	Geometric	Significance
0.2012	0.0141	4.0940×10^{-28}

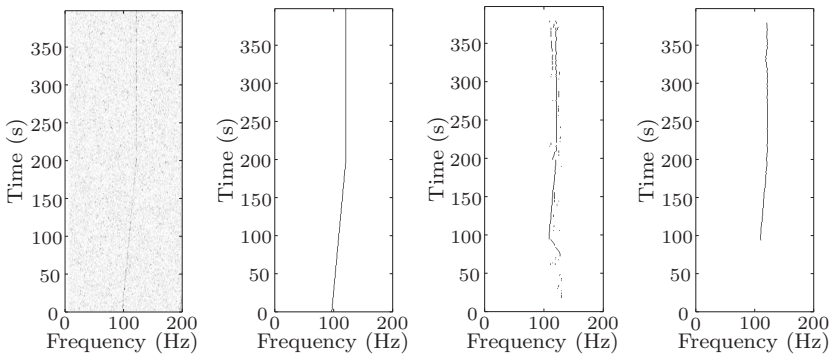
Table 2: The mean false positive detection rates obtained during evaluation on the vertical track dataset and their significance.

similar LLA (0.5515 compared to 0.5249, respectively). This means that the LLAs presented in this section should not be considered independently of the accompanying false positive rates.

When applied to the detection of oblique tracks, Fig. 11, it would appear from the LLAs that the original variant demonstrates more resilience to reduced SNRs than the geometric variant. When detecting tracks with a gradient of 1 Hz/s the algorithm variants achieve mean LLAs of 0.8996 (geometric) and 0.8728 (original) at SNRs of 5–7 dB and at 3–7 dB the averages reduce to 0.8355 (geometric) and 0.8715 (original). When detecting tracks with greater gradients, the algorithm variant achieving the best LLA performance reverses. For example, the mean values for the detection of 2 Hz/s gradient tracks in the same SNR range are: 0.4386 (geometric) and 0.4658 (original). Tracks that have a gradient of 8 Hz/s and 16 Hz/s seem beyond the capabilities of both variants and the LLAs are close to zero. It is shown in Table 3, however, that once again the false positive rates of the geometric variant are much lower: the maximum being 0.0035, whereas the maximum obtained using the original variant is 0.0623. Therefore the LLAs of the geometric variant are far more reliable and represent better detection results.

The detection of sinusoidal tracks (Figs. 12, 13 and 14) results in a similar outcome. As the amplitude (the centre frequency variation) of the sinusoid increases the LLAs of the detections decrease, which holds for both variants of the algorithm. Furthermore, at first sight the original variant of the algorithm seems more resilient to reduced SNRs. When applied to the detection of sinusoidal tracks with a period of ten and fifteen seconds (Fig. 12 and Fig. 13) and at SNRs greater than 4 dB both variants result in very similar LLAs. When, however, the algorithms are applied to the detection of sinusoidal tracks with a period of twenty seconds (Fig. 14) the geometric variant produces better LLAs at SNRs above 4 dB.

The false positive rates presented in Tables 4, 5 and 6 give more insight into these LLA results. They



(a) Spectrogram (b) Ground Truth (c) Original (d) Geometric.

Figure 10: Example detections (note that the top and bottom twenty pixels of the spectrogram are not considered due to a combination of the length of the contour, k, and the height of the potential energy window, N'), in which the LLAs are 0.5249 (original) and 0.5515 (geometric); the mean false positive rates are 0.4302 (original) and 0.0000 (geometric); and the SNR of the spectrogram is 5.5 dB.

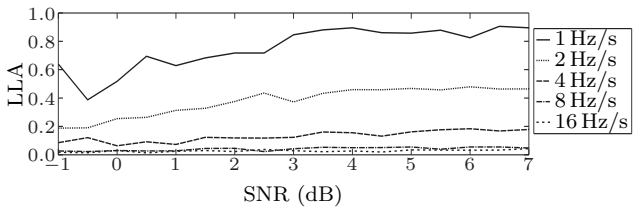
	Original	**Geometric**	**Significance**
1 Hz/s	0.0623	0.0011	1.7347×10^{-24}
2 Hz/s	0.0441	0.0035	3.6942×10^{-18}
4 Hz/s	0.0189	0.0007	2.9693×10^{-15}
8 Hz/s	0.0142	0.0010	1.9388×10^{-17}
16 Hz/s	0.0141	0.0015	2.2420×10^{-15}

Table 3: The mean false positive detection rates obtained during evaluation on the oblique track dataset and their significance.

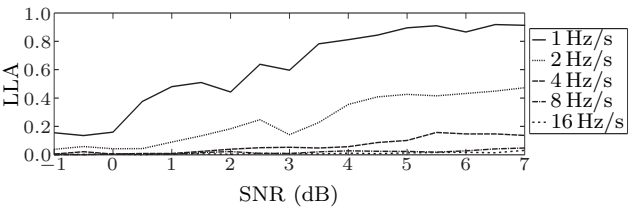
demonstrate that once again the original variant produces many more false positive detections than the geometric variant, which results in a maximum of 0.1029 false positive detections per row. The original variant produces 0.4339. For the remaining cases the mean probability of false positives per row is largely between 0.02–0.03 for the geometric variant and 0.2–0.3 for the original—the geometric variant producing a factor of ten less than the original variant.

3.4.3. Discussion

Viewing the LLAs alone would give the belief that both of these algorithms, particularly at high signal-to-noise ratios, are evenly matched. When, however, the probability of false positive detections per row is taken into account, the geometric variant provides far more favourable results across all the experiments. In terms of LLA, the original algorithm has shown more resilience to decreasing SNRs, however, this is attributed to the increased number of false positive detections, which artificially inflate its LLA. The geometric energy formulation results in a mean probability of false positive detections per row of less than 0.015, 0.004, 0.110, 0.030 and 0.030 in the vertical, oblique, ten second sinusoidal, fifteen second sinusoidal and twenty second

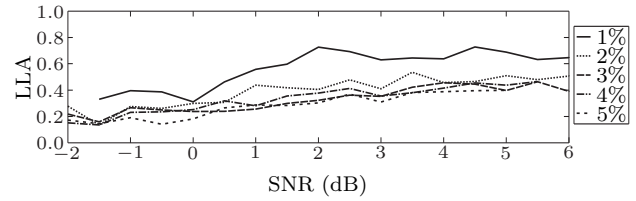


(a) Original internal energy algorithm variant. The mean standard deviations for the each function is: 1 Hz/s = 0.0128, 2 Hz/s = 0.0075, 4 Hz/s = 0.0052, 8 Hz/s = 0.0031 and 16 Hz/s = 0.0031.

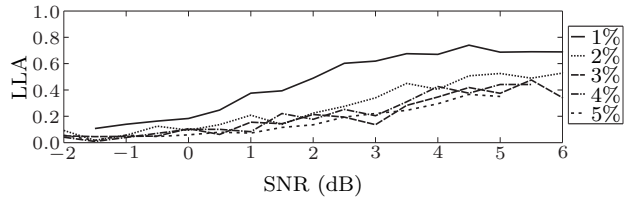


(b) Geometric internal energy algorithm variant. The mean standard deviations for the each function is: 1 Hz/s = 0.0126, 2 Hz/s = 0.0098, 4 Hz/s = 0.0053, 8 Hz/s = 0.0029 and 16 Hz/s = 0.0015.

Figure 11: The mean line location accuracies of oblique track detections as functions of the spectrogram's SNR.



(a) Original internal energy algorithm variant. The mean standard deviations for the each function is: 1% = 0.0178, 2% = 0.0112, 3% = 0.0203, 4% = 0.0166 and 5% = 0.0205.



(b) Geometric internal energy algorithm variant. The mean standard deviations for the each function is: 1% = 0.0044, 2% = 0.0056, 3% = 0.0083, 4% = 0.0166 and 5% = 0.0095.

Figure 12: The mean line location accuracies of sinusoidal (ten second period) track detections as functions of the spectrogram's SNR.

sinusoidal track experiments respectively. These are low in comparison to the original variant, which results in maximum false positive rates of approximately 0.201, 0.062, 0.434, 0.267 and 0.297.

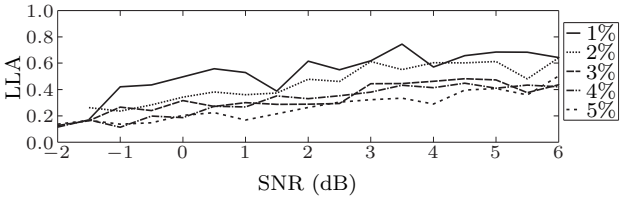
It can be seen that, irrespective of the internal energy, the active contour algorithm is able to detect all variations of the track structure. There are limitations to this, however; tracks that have a gradient greater than 4 Hz/s are beyond the deformable capabilities of the contour or the generalisation capabilities of the potential energy. Sinusoidal tracks are detected with a high probability at SNRs above 3.5 dB, and the same can be said for the detection of oblique tracks. Vertical tracks are reliably detected at very low SNRs of around −1 dB.

3.5. Real Data

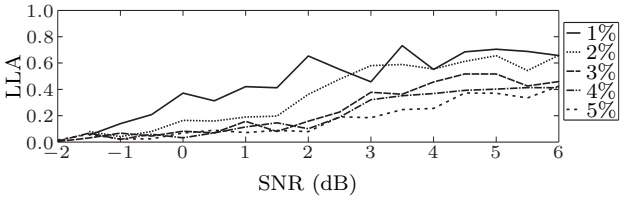
The algorithm was applied to the example of real data presented in Fig. 15a, an underwater recording of a fishing vessel, and the result is presented in Fig. 15b. The method successfully detects the track present in the spectrogram. The detection contains gaps where the track fades as there are no harmonics present to

	Original	Geometric	Significance
1%	0.2225	0.0189	3.8444×10^{-26}
2%	0.2778	0.0334	5.8467×10^{-25}
4%	0.4017	0.0827	2.1695×10^{-22}
8%	0.4339	0.1029	4.7220×10^{-21}
16%	0.4266	0.0780	2.6199×10^{-22}

Table 4: The mean false positive detection rates obtained during evaluation on the sinusoidal track dataset (10 second period) and their significance.



(a) Original internal energy algorithm variant. The mean standard deviations for the each function is: 1% = 0.0163, 2% = 0.0099, 3% = 0.0148, 4% = 0.0137 and 5% = 0.0101.



(b) Geometric internal energy algorithm variant. The mean standard deviations for the each function is: 1% = 0.0051, 2% = 0.0055, 3% = 0.0049, 4% = 0.0050 and 5% = 0.0045.

Figure 13: The mean line location accuracies of sinusoidal (fifteen second period) track detections as functions of the spectrogram's SNR.

augment the missing information (which would be the case in multiple track detection). It also demonstrates the method's ability to model small fluctuations in the frequency position of the track.

3.6. Further Discussion

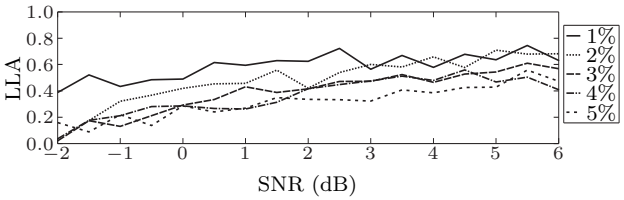
After analysing each experiment in detail it is now possible to discuss some general findings with respect to the proposed algorithm and the LLA measure.

3.6.1. The Proposed Algorithm

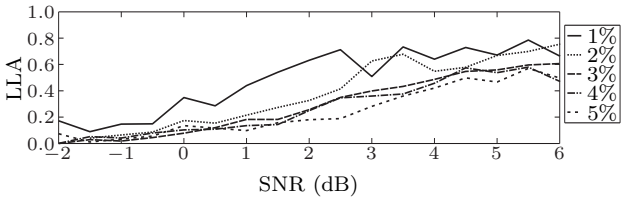
It has been shown during optimisation that α has little influence on the detection performance. This was observed in two variants of the algorithm (using the original and proposed potential energy). The parameter β behaves similarly in both variants, the key difference is that the LLA maximum is reached at a higher parameter value in the original variant. The empirical evidence indicates that the geometric internal energy captures more salient information, and therefore its influence does not need to be as strong; the geometric variant achieves similar or higher LLAs compared with the original variant whilst having a lower influence on the contour's energy. Generally, high weighting of the internal energies restricts the contour's ability to deform and evolve, resulting in a large degradation in performance. The gradient potential has the characteristics of a threshold: a high value reduces false positive detections and concomitantly reduces

	Original	Geometric	Significance
1%	0.2483	0.0242	7.0833×10^{-25}
2%	0.2343	0.0269	1.5069×10^{-24}
4%	0.2669	0.0291	9.6980×10^{-32}
8%	0.2564	0.0264	1.7339×10^{-29}
16%	0.2346	0.0233	8.9775×10^{-28}

Table 5: The mean false positive detection rates obtained during evaluation on the sinusoidal track dataset (15 second period) and their significance.



(a) Original internal energy algorithm variant. The mean standard deviations for the each function is: 1% = 0.0107, 2% = 0.0068, 3% = 0.0171, 4% = 0.0183 and 5% = 0.0121.



(b) Geometric internal energy algorithm variant. The mean standard deviations for the each function is: 1% = 0.0046, 2% = 0.0042, 3% = 0.0061, 4% = 0.0064 and 5% = 0.0057.

Figure 14: The mean line location accuracies of sinusoidal (twenty second period) track detections as functions of the spectrogram's SNR.

true positive detections, a low value increases false positive detections—a point of attention when selecting its value.

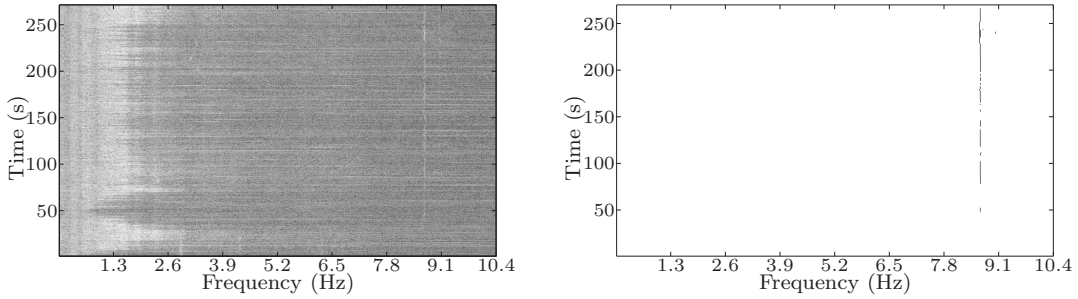
The algorithm introduces the capability of detecting multiple tracks simultaneously. Integrating harmonic positions reduces the response to false positive detections (as noise is not harmonic), allowing for a reduction in the gradient potential force, and creating a more sensitive detection strategy. The proposed potential energy produces far fewer false positive detections when compared with the original. Increasing its influence produces a stable probability of false positive detections, instead of increasing them, as with the original. Principally, the proposed algorithm, has proven to be a reliable method for extracting unknown shaped tracks in spectrograms.

3.6.2. Line Location Accuracy

The LLA measure aggregates measures of the true positive rate; the false positive rate; and the location of detection. Optimisation using LLA forms a balance between the three criteria. This balance, however, does not necessarily lead to a well performing algorithm. It can, in fact, mask weak detection performance, ranking it similar to that of a good detector. Accompanying the LLA with false positive detection rates allows these two cases to be disambiguated. Nevertheless, depending upon the application some criteria may outweigh others. For example, it may be sufficient to identify the presence of a source in a particular

	Original	Geometric	Significance
1%	0.2003	0.0171	1.3361×10^{-30}
2%	0.2006	0.0179	3.3317×10^{-28}
4%	0.2482	0.0313	1.8298×10^{-27}
8%	0.2973	0.0332	5.4875×10^{-31}
16%	0.2798	0.0299	5.1713×10^{-28}

Table 6: The mean false positive detection rates obtained during evaluation on the sinusoidal track dataset (20 second period) and their significance.



(a) An example of a real spectrogram image where intensity represents $\log(\mathbf{S})$.

(b) The detection resulting from using the active contour algorithm.

Figure 15: An example of real spectrogram track detection. Parameters took the following values: $c = 0.17$, $\alpha = 0.10$, $\beta = 0.20$, $N' = 21$, $M' = 3$, and $k = 21$.

time frame and knowledge its exact frequency may be superfluous. In this case, it may be more desirable to maximise true-positive detections or minimise false-positive detections at the expense of other criteria.

4. Conclusions

This paper has presented a novel algorithm for the detection of single and multiple tracks in spectrograms by, optionally, integrating information from harmonically related positions. A discussion of the original active contour algorithm, its limitations and alternative methodologies, has also been presented. This has led to the identification of issues that prevent the algorithm in its original form from being applied to spectrogram track detection. Novel solutions to these problems have been proposed. A gradient potential has been proposed to overcome the dependency on initialisation position. A novel potential energy formulation has been proposed to take advantage of structural and intensity information to increase detection rates. The undesirable bias resulting from the original internal energies has been removed by replacing them with the geometric internal energy, which has been shown to achieve higher LLAs and reduce false positive detections. An analysis of the algorithm's computational complexity has shown that it is applicable to real-time implementations. Moreover, the findings presented within this paper have been related to existing work in the area and

weaknesses of the LLA measure have been revealed.

Acknowledgments

This research has been supported by the DSTL[1] and QinetiQ Ltd.[2], with special thanks to Duncan Williams[1] and Jim Nicholson[2]. We are grateful to Simon Poulding for his advice on statistical evaluation.

Bibliography

[1] Abel, J. S., Lee, H. J., Lowell, A. P., March 1992. An image processing approach to frequency tracking. In: Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Process. Vol. 2. pp. 561–564.

[2] Altes, R. A., April 1980. Detection, estimation, and classification with spectrograms. Journal of the Acoustical Society of America 67 (4), 1232–1246.

[3] Ballerini, L., May 1999. Genetic snakes for medical images segmentation. In: Proc. of the 1st Eur. Workshop on Evol. Image Anal., Signal Process. & Telecommunications. Vol. 1596 of LNCS. pp. 59–73.

[4] Bishop, C. M., 1995. Neural Networks for Pattern Recognition. Oxford University Press Inc.

[5] Brent, R. P., 1976. Fast multiple-precision evaluation of elementary functions. J. ACM 23, 242–251.

[6] Brillinger, D. R., 1975. Time Series: data analysis and theory. International Series in Decision Processes. Holt, Reinhart and Winston Inc., New York.

[7] Chalana, V., Linker, D. T., Haynor, D. R., Kim, Y., 1996. A multiple active contour model for cardiac boundary detection on echocardiographic sequences. IEEE Trans. on Med. Imag. 15 (3), 290–298.

[8] Chan, T. F., Vese, L. A., February 2001. Active contours without edges. IEEE Transactions on Image Processing 10 (2), 266–277.

[9] Cohen, L. D., March 1991. On active contour models and balloons. Computer Vision, Graphics, and Image Processing: Image Understanding 53 (2), 211–218.

[10] Cormen, T. H., Leiserson, C. E., Rivest, R. L., Stein, C., September 2001. Introduction to Algorithms, 2nd Edition. MIT Press.

[11] Dilengette, H., 1994. Intrinsic Stabilizers of Planar Curves. In: Proc. of the Third European Conf. on Computer Vision. pp. 427–436.

[12] Di Martino, J.-C., Colnet, B., Di Martino, M., 1994. The use of non supervised neural networks to detect lines in lofargram. In: Proc. of the Int. Conf. on Acoust. Speech & Signal Process. pp. 293–296.

[13] Di Martino, J.-C., Haton, J. P., Laporte, A., April 1993. Lofargram line tracking by multistage decision process. In: Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Process. pp. 317–320.

[14] Di Martino, J.-C., Tabbone, S., September 1995. Detection of lofar lines. In: Proc. of the 8th Int. Conf. on Image Anal. & Process.V̇ol. 974 of LNCS. pp. 709–714.

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

[16] Gillespie, D., 2004. Detection and classification of right whale calls using an 'edge' detector operating on a smoothed spectrogram. Canadian Acoustics 32 (2), 39–47.

[17] Grigorakis, A., August 1997. Application of detection theory to the measurement of the minimum detectable signal for a sinusoid in Gaussian noise displayed on a lofargram. Tech. Rep. DSTO-TR-0568, Maritime Operations Division, Aeronautical and Maritime Research Laboratory, Defence Science and Technology Organisation, Canberra.

[18] Ji, L., Yan, H., April 2002. Attractable snakes based on the greedy algorithm for contour extraction. Pattern Recognition 35 (4), 791–806.

[19] Kass, M., Witkin, A., Terzopoulos, D., January 1988. Snakes: Active contour models. International Journal of Computer Vision 1 (4), 321–331.

[20] Kendall, G. D., Hall, T. J., Newton, T. J., June 1993. An investigation of the generalisation performance of neural networks applied to lofargram classification. Neural Computing & Applications 1 (2), 147–159.

[21] Koenig, W., Dunn, H. K., Lacy, L. Y., July 1946. The sound spectrograph. Journal of the Acoustical Society America 18 (1), 244–244.

[22] Kootsookos, P. J., 1993. A review of the frequency estimation and tracking problems. Tech. rep., Systems Engineering Department, Australian National University.

[23] Lampert, T. A., O'Keefe, S. E. M., July 2010. An active contour algorithm for spectrogram track detection. Pattern Recognition Letters 31 (10), 1201–1206.

[24] Lampert, T. A., O'Keefe, S. E. M., September 2011. A detailed investigation into low-level feature detection in spectrogram images. Submitted to Pattern Recognition. Pattern Recognition 44 (9), 2076–2092.

[25] Lampert, T. A., O'Keefe, S. E. M., February 2010. A survey of spectrogram track detection algorithms. Applied Acoustics 71 (2), 87–100.

[26] Morrissey, R. P., Ward, J., DiMarzio, N., Jarvis, S., Moretti, D. J., November–December 2006. Passive acoustic detection and localisation of sperm whales (*Physeter Macrocephalus*) in the tongue of the ocean. Applied Acoustics 67 (11–12), 1091–1105.

[27] Paris, S., Jauffret, C., March 2001. A new tracker for multiple frequency line. In: Proceedings of the IEEE Conference on Aerospace. Vol. 4. IEEE, pp. 1771–1782.

[28] Paris, S., Jauffret, C., April 2003. Frequency line tracking using HMM-based schemes. IEEE Transactions on Aerospace and Electronic Systems 39 (2), 439–450.

[29] Perrin, D. P., Smith, C. E., 2001. Rethinking classical internal forces for active contour models. In: Proceedings of the IEEE Int. Conference on Computer Vision and Pattern Recognition. pp. 615–620.

[30] Poggio, T., Torre, V., Koch, C., September 1985. Computational vision and regularization theory. Nature 317, 314–319.

[31] Poole, D., 2005. Linear Algebra: A Modern Introduction, 2nd Edition. Brooks/Cole.

[32] Pratt, W. K., 1991. Digital Image Processing, 2nd Edition. Wiley-Interscience Publication.

[33] Prewitt, J. M. S., 1970. Picture Processing and Psychopictorics. Academic Press Inc., New York, NY, USA, Ch. Object Enhancement and Extraction, pp. 75–149.

[34] Quinn, B. G., May 1994. Estimating frequency by interpolation using Fourier coefficients. IEEE Transactions on Signal Processing 42 (5), 1264–1268.

[35] Quinn, B. G., Hannan, E. J., 2001. The Estimation and Tracking of Frequency. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

[36] Rife, D. C., Boorstyn, R. R., September 1974. Single-tone parameter estimation from discrete-time observations. IEEE Transactions on Information Theory 20 (5), 591–598.

[37] Samadani, R., March 1989. Changes in connectivity in active contour models. In: Proceedings of the Workshop on Visual Motion. pp. 337–343.

[38] Scharf, L. L., Elliot, H., October 1981. Aspects of dynamic programming in signal and image processing. IEEE Transactions on Automatic Control 26 (5), 1018–1029.

[39] Shi, Y., Chang, E., April 2003. Spectrogram-based formant tracking via particle filters. In: Proceedings of the IEEE Int. Conference on Acoustics, Speech, and Signal Processing. Vol. 1. pp. I–168–I–171.

[40] Shih, F. Y., Zhang, K., February 2007. Locating object contours in complex background using improved snakes. Computer

Vision and Image Understanding 105 (2), 93–98.

[41] Tikhonov, A., 1963. Regularization of incorrectly posed problems. Soviet Mathematics Doklady 4, 1624–1627.

[42] Van Cappel, D., Alinat, P., October 1998. Frequency line extractor using multiple hidden Markov models. In: OCEANS '98 Conference Proceedings. Vol. 3. pp. 1481–1485.

[43] Williams, D. J., Shah, M., January 1992. A fast algorithm for active contours and curvature estimation. CVGIP: Image Understanding 55 (1), 14–26.

[44] Wold, H. O. A., 1963. Forecasting by the chain principle. Time Series Analysis, 471–497.

[45] Xie, X., Evans, R. J., December 1991. Multiple target tracking and multiple frequency line tracking using hidden Markov models. IEEE Transactions on Signal Processing 39 (12), 2659–2676.

[46] Yang, S., Li, Z., Wang, X., July 2002. Ship recognition via its radiated sound: the fractal based approaches. Journal of the Acoustical Society of America 11 (1), 172–177.