Pattern Recognition 44 (2011) 2076–2092

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A detailed investigation into low-level feature detection in spectrogram images

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

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

ARTICLE INFO

Article history:
Received 18 October 2010
Received in revised form
7 February 2011
Accepted 15 February 2011
Available online 20 February 2011

Keywords:
Spectrogram
Low-level feature detection
Periodic time series
Remote sensing
Line detection

ABSTRACT

Being the first stage of analysis within an image, low-level feature detection is a crucial step in the image analysis process and, as such, deserves suitable attention. This paper presents a systematic investigation into low-level feature detection in spectrogram images. The result of which is the identification of frequency tracks. Analysis of the literature identifies different strategies for accomplishing low-level feature detection. Nevertheless, the advantages and disadvantages of each are not explicitly investigated. Three model-based detection strategies are outlined, each extracting an increasing amount of information from the spectrogram, and, through ROC analysis, it is shown that at increasing levels of extraction the detection rates increase. Nevertheless, further investigation suggests that model-based detection has a limitation—it is not computationally feasible to fully evaluate the model of even a simple sinusoidal track. Therefore, alternative approaches, such as dimensionality reduction, are investigated to reduce the complex search space. It is shown that, if carefully selected, these techniques can approach the detection rates of model-based strategies that perform the same level of information extraction. The implementations used to derive the results presented within this paper are available online from http://stdetect.googlecode.com.

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. [26]. 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 [41], image processing [1,10] and expert systems [35]. 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.

In the broad sense this 'problem arises in any area of science where periodic phenomena are evident and in particular signal processing' [44]. 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 wide and include identifying and tracking marine mammals via their calls [39,36], identifying ships, torpedoes or submarines via the noise radiated by their mechanical movements such as propeller blades and machinery [52,7], distinguishing underwater events such as ice cracking [16] and earth quakes [20] from different types of source, meteor detection, speech formant tracking [47] 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 [36,39] has resulted in increased data collection which requires automated algorithms to detect calls and determine local specie population and numbers. The research presented in this paper is applicable to any area of science in which it is necessary to detect frequency components within time-series data.

---

* Corresponding author. Tel.: +44 1904 325500; fax: +44 1904 567687.
E-mail addresses: tomal@cs.york.ac.uk (T.A. Lampert), sok@cs.york.ac.uk (S.E. O'Keefe).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.02.014

A spectrogram is a visual representation of the distribution of acoustic energy across frequencies and over time, and is formally defined in [29]. The vertical axis of a spectrogram typically represents time, the horizontal axis represents the discrete frequency steps, and the amount of power detected is represented as the intensity at each time-frequency point. For a complete review of spectrogram track detection methods the reader is referred to a recently published survey of spectrogram track detection algorithms [29].

The methods presented can be reduced to, and can therefore be characterised by, their low-level feature detection mechanisms. Low-level feature detection is the first stage in the detection of any object within an image and it is therefore key to any higher level processing. For a spectrogram, this stage results in the identification of unconnected points that are likely to belong to a track, which are output in the form of another image [18]. It is found that a number of mechanisms are in use, however, there exists no systematic investigation into the advantages and disadvantages of each. Abel et al. [1], Di Martino et al. [9], Scharf and Elliot [46] and Paris and Jauffret [41], to name but a few, take the approach of detecting single-pixel instances of the tracks, therefore only intensity information can be exploited in the decision process. Methods such as those presented by Gillespie [17], Kendall et al. [25] and Lemming [34] use windows in a spectrogram to train neural network classifiers—the benefits of this, however, were not investigated and the research was probably motivated for the ability to use neural networks. In addition to intensity information, their approach allowed for information regarding the track structure to be exploited in the decision process. Nevertheless, an empirical study of the differences and detection benefits between the two approaches is still lacking. It would be expected that when intensity information degrades, such as in low signal-to-noise ratio spectrograms, the structural information will augment this deficit and thus improve detection rates.

This paper presents such a study. Firstly three low-level feature detectors are defined, each of which acts upon an increasing amount of information. These are termed 'unconstrained' detectors as they:

- perform an exhaustive search of the feature space;
- retain all of the information provided to them by the feature model;
- utilise the original, unprocessed, data.

The exhaustive search performed by these methods, however, means that they are computationally expensive and, as such, a number of 'constrained' detectors are examined. These 'constrained' detectors are characterised by one or more of the following:

- machine-learning techniques are utilised for class modelling;
- the data is transformed through dimensionality reduction;
- the data is transformed through preprocessing,

and therefore these detection techniques simplify the search space. All of the 'constrained' feature detectors evaluated derive feature vectors from within a window and they therefore act upon intensity and structural information. The 'constrained' detectors are split into two categories—data-based and model-based—to reflect the source of the training samples utilised by their supervised learning process. Finally, the performance of a model-based 'unconstrained' feature detector is compared against a model-based 'constrained' feature detector to ascertain the degree of performance divergence between the two approaches.

Furthermore, this paper presents a novel transformation that integrates information from harmonic locations within the spectrogram. This is possible due to the harmonic nature of acoustic signals and is defined with the aim of revealing the presence of an acoustic source at low signal-to-noise ratios by utilising all of the information available. The benefits of performing low-level feature detection whilst combining information from harmonic locations are shown at the end of this paper through a comparison with the detection performance achieved by the low-level feature detectors when applied to the original spectrogram.

The remainder of this paper is organised as follows: Section 2 presents the low-level detection mechanisms; these are evaluated in Section 3 and a discussion of findings is presented; and finally the conclusions of the investigation are drawn in Section 4.

2. Method

In this section several low-level feature detection mechanisms are described and investigated. By definition, the detection of lines and edges forms two distinct problems and is commonly approached differently [18]; an edge is defined by a step function, and a line by a ridge function. Edge detectors such as the Canny operator, along with more recent methods [32], are specifically defined to detect step features and are therefore not evaluated here. The Laplacian detector is, however, an edge detector which can be applied to line detection [18] and therefore it is evaluated in Section 3 of this paper.

2.1. 'Unconstrained' feature detectors

Detection methods that utilise dimensionality reduction techniques such as principal component analysis [22] to reduce the model or data complexity, lose information regarding the feature model in the process [6]. Pre-processing of the data also introduces information loss. This information loss detracts from a detector's ability to detect features and therefore they produce sub-optimal detection results. A method which models the data correctly and does not lose any information in the detection process will have the most discrimination power as a feature detector, under the condition that it correctly models the features to be detected. These types of detectors are more generally referred to as correlation methods in the image analysis domain. In order for such methods to detect features that vary greatly, a model has to be defined with parameters corresponding to each variation type that can be observed. An exhaustive search for the parameter combination that best describes the data is conducted by matching the model to the unprocessed data by varying its parameters. In this section are defined three detection methods with the properties of an 'unconstrained' feature detector, i.e. no model reduction or approximation is performed during the search for the feature, and no preprocessing of the data that may destroy information is carried out (for example filtering or calculating gradient information). Three modes of detection have been identified, each of which increases the amount of information available to the detection process from the previous mode: individual pixels; local intensity distribution; and local structural intensity distribution. Individual pixel classification performs detection based upon the intensity value of single pixels. By definition this method makes no assumption as to the track shape and consequently is the most general of the methods in terms of detecting variable structure. A track, however, 'is a spectral representation of the temporal evolution of the signal' [8] and, therefore, 'can be expressed as a function of the time' [8], i.e. it is composed of a collection of pixels in close proximity to each other. Performing the detection process using individual pixels