Multi-target pitch tracking of vibrato sources in noise using the GM-PHD filter

Dan Stowell and Mark D. Plumbley DAN.STOWELL@EECS.QMUL.AC.UK
Centre for Digital Music, Queen Mary University of London, UK

Abstract

Probabilistic approaches to tracking often use single-source Bayesian models; applying these to multi-source tasks is problematic. We apply a principled multi-object tracking implementation, the Gaussian mixture probability hypothesis density filter, to track multiple sources having fixed pitch plus vibrato. We demonstrate high-quality filtering in a synthetic experiment, and find improved tracking using a richer feature set which captures underlying dynamics. Our implementation is available as open-source Python code.

Probabilistic modelling of audio objects is useful because Bayesian methods can be used to make principled inferences about the content of audio signals. For reasons of simplicity and tractability, inferences based on single-source models are widely used, such as the standard Hidden Markov Model (HMM) approach to speech recognition and music modelling. However, music is very often polyphonic, so there is a need to analyse acoustic scenes in which multiple sources may be simultaneously active. Multi-source tracking can be achieved by repeated application of single-source models, but this does not reflect the true scene and may yield sub-optimal results (Mahler, 2007).

Existing multi-source approaches in music informatics often use non-probabilistic techniques. Probabilistic approaches exist, such as Probabilistic Latent Component Analysis (PLCA) which characterises sources as time-varying activations of spectral bases. However, such models are not always well-matched to audio objects with structured variability over time, and are poorly suited to causal (e.g. real-time) tracking.

In this paper we investigate an alternative multiple tracking paradigm, which models a set-valued random variable having multiple objects (Mahler, 2007). The *probability hypothesis density* filter (PHD filter) is one practical realisation of this approach. Given a system with linear Markov state updates and a linear observation model, it propagates a density through time which is an estimate of the underlying system state. The PHD filter was originally formulated as a particle-type filter. Later work introduced the Gaussian mixture PHD filter (GM-PHD filter), using a Gaussian mixture (GM) to represent state and having improved performance (Vo & Ma, 2006; Mahler, 2007).

The GM-PHD filter has similarities to a HMM- or Kalman-type filter with hidden state represented as a GM, propagated from one time-frame to the next. However, the GM does not represent a probability density but the "intensity": the first moment of the set-valued system state. The intensity does not integrate to 1, but to a total reflecting the expected number of objects present; its value at a location can be thought of as the expected number of objects at that location.

1. Implementation

The GM-PHD filter has been applied to music audio in one published work (Clark et al., 2007), to post-process the output of sinusoidal modelling of piano notes, using a model assuming fixed pitch and decaying amplitude. Here we explore its application to tracking multiple pitched sources in noise, where the sources may exhibit vibrato modulations that obscure the observed pitch. In particular, we wish to study whether the filter can recover stable tracks from observations such as might be found by dictionary-type approaches, whose output is often a variable number of observations.

In practice, a missed detection may be more or less desirable than a false positive, and some reweighting of the tendency to positive or negative errors is desired. In the following, we use a multiplicative bias factor to alter this tendency: we multiply the total weight by the bias factor, before rounding the result to give the

Appearing in *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, Scotland, UK, 2012.

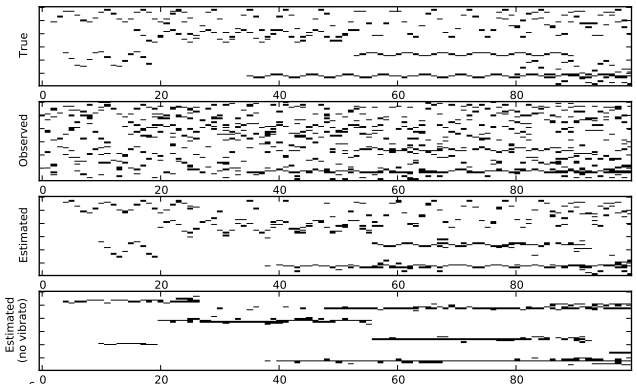

Figure 1. True state, observation, filter outputs (one trial).

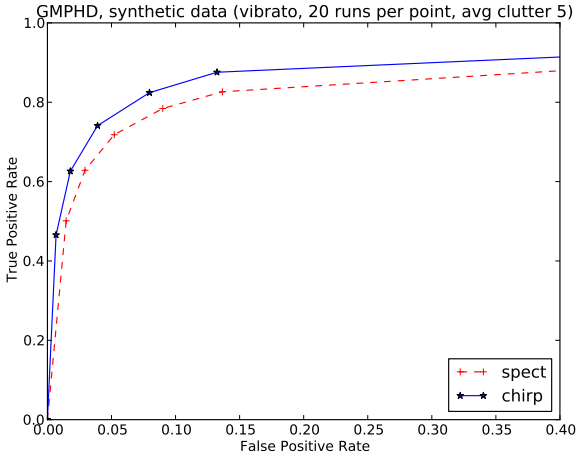

Figure 2. Filter accuracy at different bias levels. In this case, an upward bias of 4 or 8 yields approximately equal false-positive and false-negative rates.

number of states to extract. As we will see, an upward bias gives us better results in our particular scenario.

2. Experiment

Our aim is to apply multi-object tracking to audio analysed with dictionary representations, to recover traces of multiple audio objects over time. As a first step, we apply the GM-PHD filter to synthetic data of a similar format to spectral and chirplet features. We synthesise data according to the following scenario:

• Over a period of 100 frames, "notes" appear and disappear, each having a fixed random pitch (within [0, 60] units) and vibrato extent (±5 units). New notes appear once per ten frames on average; each note ends with probability 0.025 per frame.

• Each note has a 3D internal state [pitch, x, dx] where the observation is at pitch+x, and the update is such that x and dx together create the vibrato oscillation.

• Observations are either spectrum-like (each observation is a frequency value) or chirplet-like (a pair of frequency values): a note is detected with a probability of 0.95 at each frame; "clutter" observations are generated by a Poisson process averaging 5 per frame.

The GM-PHD filter receives the set-valued observations causally, and also the linear model for to the notes' vibrato evolution, which it uses to infer state.

In Figure 1 we show the results of one trial. The true state (depicted in the top panel, in the form of noiseless observations) consists of a variable number of oscillating sources; observations reflect these imperfectly, typically receiving more clutter observations than true observations. Despite this, the filter is able to reject much of the clutter. It also recovers the underlying pitch values, separating out the vibrato.

To evaluate performance, we generate binary spectrogram-like images from the true and the inferred state (as in Figure 1), then count true and false detections on a per-bin basis. We repeat this procedure at a range of different bias levels, averaging results over 20 independent trials. Figure 2 summarises the results, showing strong filter performance, with the chirplet-like representation consistently outperforming the spectrum-like representation.

3. Conclusions

We have shown in a synthetic experiment that the GM-PHD filter can track multiple simultaneous vibrato notes in noise, separating the vibrato from the underlying pitch, and that a rich feature representation improves the detection accuracy. The GM-PHD filter requires a linear physical model of the sources' evolution for tracking; future work will investigate whether this can be adapted to a wider range of evolving audio sources. Our implementation is available as open-source Python code.[1]

References

Clark, D. *et al.* Multi-object tracking of sinusoidal components in audio with the GM-PHD filter. In *Proc WASPAA*, pp. 339–342. 2007.

Mahler, R. *Statistical multisource-multitarget information fusion.* Artech House, Boston/London, 2007.

Vo, B. N. and Ma, W. K. The Gaussian mixture probability hypothesis density filter. *IEEE Transactions on Signal Processing*, 54(11):4091–4104, 2006.

[1] https://code.soundsoftware.ac.uk/projects/gmphd