ONSET DETECTION FOR MIREX 2016

Che-Yuan Liang, Li Su, and Yi-Hsuan Yang
Academia Sinica
{mister2dot4, lisu, yang}@citi.sinica.edu.tw

ABSTRACT

The submission presents an onset detector using constrained linear reconstruction, as an extension to spectral flux methods. Instead of comparing only the spectral content of every consecutive two frames, the proposed method takes into account a wider temporal context to evaluate the dissimilarity between a given frame and its previous frames. The proposed method is efficient, easy to implement, and is applicable to scenarios of online onset detection.

1. INTRODUCTION

The main idea of this submission is to extend the spectral flux method by taking into account more temporal information while formulating the onset detection function (ODF). Specifically, while the spectral flux method measures the "audio novelty" (i.e. value in the ODF) of a given frame by comparing it with one of its previous frames, the proposed method measures the audio novelty by the reconstruction error while we use the linear combination of a collection of the previous frames to approximate the given frame. In other words, we assume that an onset event is the time instance when the current frame cannot be easily predicted by its previous frames. In this way, the computation of the ODF becomes a convex optimization problem, where additional constraints, such as the sparsity or non-negativity of the combination coefficients, can be added. Detail information of this method can be found in [3].

2. GENERALIZED ONSET DETECTION FUNCTION

Instead of merely comparing the spectral difference between two frames, the proposed method considers up to τ previous frames [3] and with a temporal lag of μ frames [1]. We formulate the calculation of the ODF as the following linear reconstruction problem:

$$\begin{aligned} \{\mathbf{r}_n^*, \boldsymbol{\alpha}_n^*\} &= \underset{\boldsymbol{r}_n, \boldsymbol{\alpha}_n}{\arg\min} \|\mathbf{r}_n\|_2^2 + \lambda \cdot g(\boldsymbol{\alpha}_n)\,, \\ \mathbf{r}_n &= \bar{\mathbf{x}}_n - \bar{\mathbf{X}}_n \boldsymbol{\alpha}_n\,, \end{aligned} \tag{1}$$

where $\boldsymbol{\alpha}_n \in \mathbb{R}^\tau$ denotes the combination coefficients of the τ previous frames $\bar{\mathbf{X}}_n \in \mathbb{R}^{K\times\tau}$ to reconstruct the current frame, $\|\mathbf{z}\|_2^2 = \sum_{k=1}^K z_k^2$ is the squared l_2 norm, and $g(\cdot)$ denotes the regularization term penalized by λ. To solve (1), we require that all the input feature vectors to be l_2-normalized beforehand [2,4]. More specifically, $\bar{\mathbf{x}}_n = \mathbf{x}_n/\|\mathbf{x}_n\|_2$ and $\bar{\mathbf{X}}_n = [\bar{\mathbf{x}}_{n-\mu},\ \bar{\mathbf{x}}_{n-\mu-1},\ \dots\ \bar{\mathbf{x}}_{n-\mu-(\tau-1)}]$. The parameter τ, referred to as the *reconstruction length*, is a non-negative integer. Moreover, the residual $\|\mathbf{r}_n\|_2 = \|\bar{\mathbf{x}}_n - \bar{\mathbf{X}}_n\boldsymbol{\alpha}_n\|_2$ is viewed as the reconstruction error and is expected to be indicative of onset events.

3. NNLS ONSET DETECTION ALGORITHM

In both submission LSY1 and LSY2. we consider the implementations of (1) with *non-negative least square* (NNLS), by setting $\lambda \to \infty$ and $g(\boldsymbol{\alpha}_n) = \sum_{i=1}^{\tau}(\alpha_{ni})_-$, where α_{ni} denotes the i-th element of $\boldsymbol{\alpha}_n$ and the function $(z)_-$ returns 1 if $z < 0$ and 0 otherwise. In other words, $g(\boldsymbol{\alpha}_n) = 0$ if and only if all the elements in $\boldsymbol{\alpha}_n$ are non-negative. With this constraint, we ensure that the reconstruction is only additive. The ODF is formulated as below:

$$\mathrm{ODF}_{LR}(n) = \|\mathbf{r}_n \odot (\mathbf{x}_n - \mathbf{x}_{n-\mu})_+\|_2 \cdot \|\mathbf{x}_n\|_2\,, \tag{2}$$

where $\odot$ denotes the element-wise product and $+$ is the rectification operation derived from the un-normolized spectrum. Only the frequency bands with increased energy from $n-\mu$ to n in the original, un-normalized feature vectors are considered in calculating the reconstruction error.

For the input feature $\bar{\mathbf{X}}_n$, we adopt the spectral feature designed in SuperFlux, which has been found effective in reducing the number of false positives originating from vibrato, and has become the input feature of the state-of-the-art onset detection algorithms in MIREX 2013 and 2014 [1]. Implementation details of the proposed method can be found in [3].

To discriminate with the submission of LSY in MIREX 2015, LSY1 has an updated peak-picking threshold parameter and LSY2 has an updated τ with longer temporal context for the linear reconstruction.

4. SOURCE CODE

The source code of the submission is available online `https://github.com/cheyuanl/OnsetDetectorCLR`
Our implementation is based on Superflux [1].

5. REFERENCES

[1] Sebastian Böck and Gerhard Widmer. Maximum filter vibrato suppression for onset detection. In *Proc. Int. Conf. Digital Audio Effects*, 2013. [Online] `https://github.com/CPJKU/SuperFlux`.

[2] Patrik O Hoyer. Non-negative sparse coding. In *Proc. IEEE Workshop Neural Networks for Signal Processing*, pages 557–565, 2002.

[3] Che-Yuan Liang, Li Su, and Yi-Hsuan Yang. Musical onset detection using constrained linear reconstruction. *IEEE Signal Processing Letter*, 22(11):2142–2146, 2015.

[4] Donald W Marquardt. Comment: You should standardize the predictor variables in your regression models. *J. American Statistical Association*, 75(369):87–91, 1980.