IMPROVED PRE-FILTERING TECHNIQUE

Chirag Mohapatra[1] and Tuhin Girinath[1]

[1] Indian Institute of Technology , Delhi

Abstract— **The basic idea proposed here is to develop a pre-filtering technique more efficient than the current available ones using principles of Convolution and repetition .**

I. INTRODUCTION

A musical chord is defined as a group of notes sounding either simultaneously or in close succession. Chords are one of the smallest and most fundamental building blocks of the tonal system, arguably the most effective means of representing Western popular music. A chord progression (a sequence of chords) outlines the harmonic flow of a piece of music. As chords tend to change on musical beats, a chord progression can represent not only the harmonic content but also the rhythmic properties of a song (e.g., the harmonic rhythm). In addition, the repeating chord patterns in a chord progression provide information useful for determining the structure or musical form of a song. Furthermore, chord transcriptions can capture many important aspects of music in compact form, and have therefore attracted attention as a convenient and useful resource for a wide variety of music

The original motivation of this project is for it to serve as a precursor to develop a better method of chord recognition .

Determining chord boundaries is very important in chord transcription . However , the chroma features are suspectible to noise and other local variables . Filtering stage is necessary to mitigate the effect of these local vibrations .

Pre-filtering is a technique which is applied directly to chromagrams prior to the pattern matching process . Some existing techniques are the moving average and the median filters . For L > 1 consecutive frames , the moving average and the median fiters are calculated by taking the mean/median of the consecutive frames . They are defined as :

Moving average filter :

$$\bar{C}(c,m) = \frac{1}{L} \sum_{i=0}^{L-1} C(c, m+i-[(L-1)/2])$$

Median filter :

$$\bar{C}(c,m) = median[C(c,i), C(c,i+1).....C(c,j)]$$

where i = m - [(L-1)/2] and j = m + [(L-1)/2]

However , the conventional methods of pre-filtering lead to blurring of the chord boundaries and can lower the performance of post-filtering methods . The aim is to develop a pre-filtering technique which can be used in conjunction with post-filtering Vitterbi decoder for chord recognition .

II. OUR APPROACH

We suggest an approach of pre-filtering which aims to reduce noise in music signals without losing the finer features or blurring out the chord boundaries . This filter is made to be used on music signals for the purpose of chord recognition by reconstructing the chroma features for chord estimation .

This method is inspired by the one proposed by Mauch, Noland, and Dixon (2009). In their approach, the information about the repetitive structure of songs is used to enhance chroma features for chord estimation. Their method is based on a conventional frame-by-frame self-similarity matrix generated from a beatsynchronous chromagram. From the matrix, they extract repeated chord progressions by examining all diagonal lines. The chord progressions are considered as repeated only when they are of equal length. The beat and bar information estimated from a song plays a crucial role in their greedy algorithm to search for repeated sections in the matrix. The found segments are merged into larger non-overlapping segment types (e.g., verse and chorus). New features are then obtained by averaging chroma features from multiple occurrences of the same segment type.

Unlike the method proposed by Mauch , the method proposed here decides which parts to include in the smoothening process by a simple thresholding operation using the technique of recurrence plots . Here the repeated sections are not limited to just large units like chorse and verse but instead include smaller units like chords .
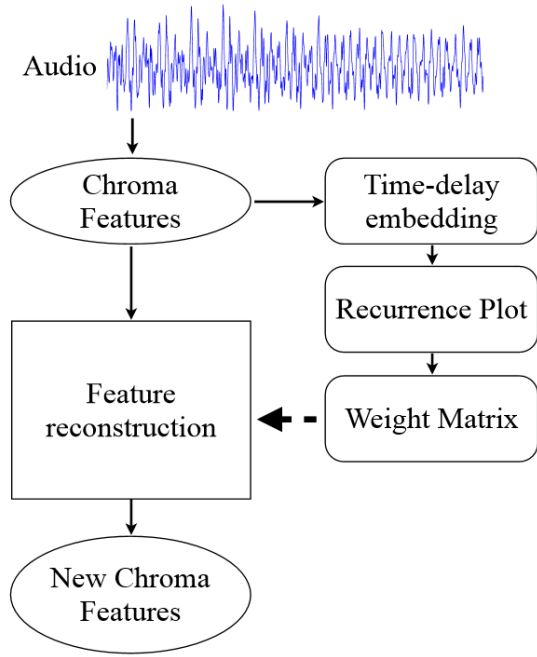
The block diagram of the proposed feature smoothing process is shown above. First, the audio signal is segmented and transformed into chroma features. The chroma features are then projected into phase space using time-delay embedding prior to calculating the recurrence plot. A weight matrix is derived from the recurrence plot, and combined with the original chroma features as a coefficient set in the feature reconstruction process .

The weight matrix is computed using recurrence plot theory, which provides a sophisticated way to analyze sequential data (Marwan, Carmen Romano, Thiel, Kurths, 2007). Recurrence plots are binary matrices representing patterns of repetition in sequential dat .. They have been previously used with chroma features in other MIR tasks such as cover version identification (Serrà, Serra, Andrzejak, 2009) and in structural similarity analysis (J. Bello, 2011). A key feature of recurrence plots is the use of time-delay embedding. Time-delay embedding is a method for transforming a time series into a multidimensional sequence of lagged data. In other words, it provides a way to transform frame-by-frame analysis into n-gram analysis (i.e., subsequence-by-subsequence).

The initial chroma features are represented in the form of a two dimensional matrix C with 12 rows(the number of chords that can be identified) and multiple columns which are a measure of how long the sequence is .

The first task is to construct the nth time-delay embedded chroma vector V(n) .

Assume that τ is the delay parameter and M is the required dimension of the embedded vector . V(n) can be constructed by concatenating all the elements of chroma sequence C(c,n) , $c \in [0,11]$ from time n to n + (M-1)t as:

V(n) = $\langle C(0,n), C(0,n+\tau)....C(0,n+(M-1)\tau)..$
$....C(11,n), C(11,n+\tau)....C(11,n+(M-1)\tau)\rangle$

After this V(n) is normalised to have unit l^2 norm .

After this , we calculate the self-similarity matrix for the embedded vector sequence . S(i,j) is representative of the relation between V(i) and V(j) . The self-similarity matrix is calculated as :

$$S(i,j) = \frac{\|V(i) - V(j)\|}{2}, \forall i, j \in [1,\hat{N}]$$

where $\hat{N}$ is the length of the time-delay embedded sequence and $\|.\|$ is the Euclidean norm . The difference is divided by 2 which is the maximum distance between two unit vectors . Hence , $0 \leq S(i,j) \leq 1 \,\forall\, i,j \in [1,\hat{N}]$

A recurrence plot with suitable threshold can be obtained from S(i,j) as :

$$R(i,j) = H[\varepsilon - S(i,j)]$$

where H is the Heaviside step function which is 0 for h < 0 and 1 otherwise .

The choice of ε is important because it is the only criterion to determine which parts are actually repeated. However, a global thresholding with a fixed value is not appropriate in this case, because the useful range of thresholds can vary greatly between songs or even within a given song.

$\varepsilon(n)$ is defined as a threshold to ensure that R(i,n) = 1 for the θ points closest to the nth point of the trajectory. In practice, we expand this approach to both columns and rows of R(i,j) to include every possible repeated pattern in the smoothing process as:

$$R(i,j) = H[\varepsilon(i) - S(i,j)] \vee H[\varepsilon(j) - S(i,j)]$$

where $\vee$ is the logical OR operator . The algorithm for computing $\varepsilon(n)$ can be described as :

Input : Self-similarity NxN matrix S(i,j)

for n ← 1 to N do

s ← nth column of S
sort s in ascending order
$\varepsilon(n) \leftarrow s(\theta)$

Then , the weight matrix is calculated as :

$$W(i,j) = R(i,j)(1 - S(i,j)) \forall i,j \in [1,\hat{N}]$$

Then , finally the smoothed chromagram is calculated by using convolution of the weight matrix and the original chromagram as :

$$\hat{C}(c,n) = \sum_{m=0}^{M-1} \frac{\sum_{i=1}^{\hat{N}} W(i,n-m).C(c,i)}{\sum_{i=1}^{\hat{N}} W(i,n-m)}$$

where $\hat{N}$ = N - M + 1 is the dimension of W , and the denominator is a normalization factor that adjusts for the contribution of overlapping chroma segments since we consider all the columns from 0 to (M-1) .

As sir suggested in the mid-term evaluation , we would look into circular convolution instead of linear convolution while calculating $\hat{C}$. The new chromagram calculated is basically a replication of all the features which were repeating in the original chromagram as indicated in the Recurrence plot . It follows the hypothesis that repetition is the basis of Music and the new chromagram can be used for identifying the chords present in the original music pattern .
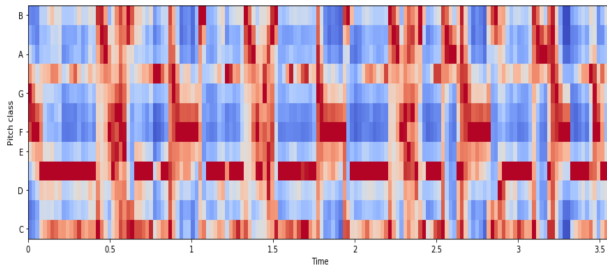

Fig:2

Fig:3

Fig:4

III. RESULTS

Fig:1

Fig:1 shows a chromagram for a particular audio file which can be accessed through this link Audio File
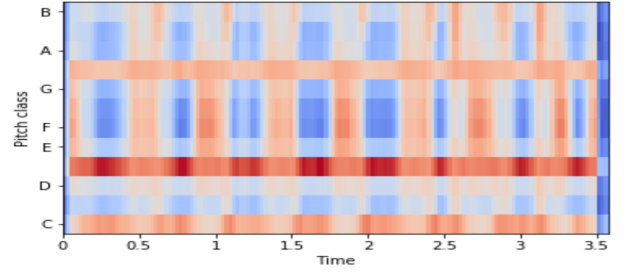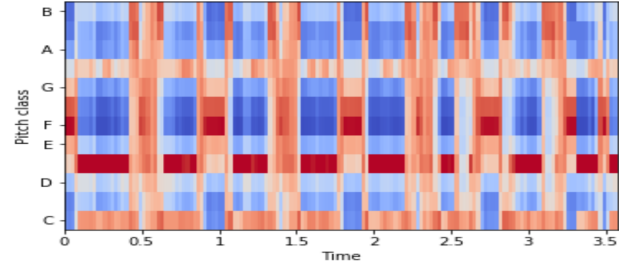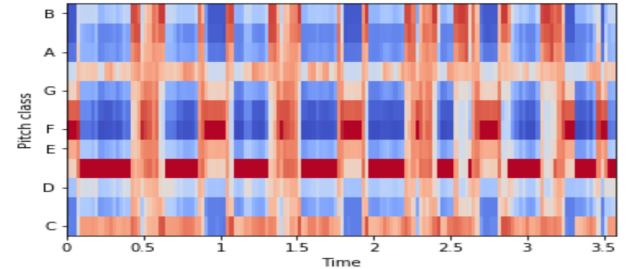
Fig:2 shows the smoothed version of the chromagram generated by using the above proposed method. Fig:3 and Fig:4 show the effect of mean and median filter on the original chromagram respectively.The chromagram generated by the proposed method is much cleaner than the original chromagram and works better for chord detection.

REFERENCES

Bello, J. (2011). Measuring structural similarity in music. IEEE Transactions on Audio, Speech, and Language Processing, 19(7), 2013-2025. 83, 85, 92

Taemin Chao (2013) . Ph.d thesis on Improved techniques for automatic chord recognition (Doctor of Philosophy in the Steinhardt School of Culture , Education and Human

Development , New York University)

Mauch, M., Dixon, S. (2010b). Simultaneous estimation of chords and musical context from audio. IEEE Transactions on Audio, Speech, and Language Processing, 18(6), 1280-1289. 35

Mauch, M., Dixon, S., Mary, Q. (2008, September). A discrete mixture model for chord labelling. In Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR) (p. 45-50). Philadelphia USA. 80

Mauch, M., Noland, K., Dixon, S. (2009, October). Using musical structure to enhance automatic chord transcription. In Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR) (p. 231- 236). Kobe, Japan. 81