ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

A Survey on Content-Based Audio Retrieval Using Chord Progression

Priyanka Gaonkar, Satishkumar Varma, Rupali Nikhare

Department of Computer Engineering, PIIT, New Panvel, University of Mumbai, India

ABSTRACT: Music information retrieval (MIR) is a science of retrieving information from music signal. MIR is a critical and challenging research topic, especially for real-time online search of similar songs over internet. Accurate and compact representation of music signals is a key component of large content-based music information retrieval (CBMIR). Here we work on how to index, and quickly and reliably retrieve relevant songs from a large-scale dataset of music audio tracks according to melody similarity. The proposed system involves compact representation of audio tracks by exploiting music content. It is done by recognizing and extracting chord progressions. The chord progressions are recognized from music signals based on a supervised statistical learning model. The extraction of chord progressions is based on support vector machine (SVM) and Hidden Markov Models (HMM). Here a chord progression histogram (CPH) is computed from each audio track as a mid-level feature, which retains the discriminative capability in describing audio content. Further the efficient organization of audio tracks is done according to their CPHs by using hash table. A set of dominant chord progressions (CPs) of each song is used as the hash key. Chord progressions and key information can serve as a robust mid-level representation and indexing for a variety of MIR tasks. A system is developed to measure the retrieval performance using precision and recall for a given query from a set of relevant music documents.

KEYWORDS*:*AudioRetrieval, Feature Extraction, chordprogression, locality sensitive hashing.

I. INTRODUCTION

Content Based Audio Retrieval (CBAR) is one of the most popular research areas on social web sites but it is critical research topic. On internet there are many music soundtracks, with the same or similar melody but sung and recorded by different people. It is very complex to find out the particular song from the multiple copies of music files having same name. For that we are developing mechanism to quickly and reliably retrieve relevant songs from a large-scale dataset of music audio tracks according to melody similarity. A melody is a linear succession of music tones. CBAR, in terms of melody similarity, has several applications such as near duplicate audio detection, relevant song retrieval and recommendation, etc. In typical scenarios, a user can find audio tracks similar to his favorite melody using an audio example, or music companies can recommend to users new music albums with similar melodies according to listening records. These applications need large-scale CBMIR techniques. Music signals usually are described by sequences of low-level features such as short-time Fourier transform (STFT), pitch, Mel frequency cepstral coefficient (MFCC), and chroma. Unfortunately, among most existing work, music audio content analysis and summarizations, by using these low-level features, are inefficient and inflexible for a scalable music information retrieval (MIR) task. In comparison, mid-level features (chord, rhythm, instrumentation) represented as musical attributes are able to better extract music structures from complex audio signals and retain semantic similarity. First we generate an accurate summary from a music signal based on chord progression. Chord is group of three or more musical notes. A chord progression is a sequence of musical chords. A chord sequence contains rich music information related to tonality and harmony, which is helpful for effectively distinguishing whether music audio tracks are similar to each other or not. Therefore, they are able to effectively and efficiently assist content-based music matching and retrieval.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

II. LITERATURE SERVEY

In this section we cite the relevant past literature that use the various chord recognition methods which are used for Music information retrieval (MIR) task. Most of the researchers concentrate on different methods for extracting feature from audio signal. These feature vectors is then used for chord recognition process.

The paper [1] present N-gram based chord recognition method for Music classification and retrieval which is proposed by Heng-Tze Cheng. They also investigated mid-level music feature construction and its applications to music classification and retrieval with recognition accuracy as competitive as existing systems. Simplicity and time-efficiency are advantages of this system. With these mid-level music features, this system able to achieve good improvement over existing approaches that use only low level features for emotional valence prediction. Limitation is failure to recognize triads differed at the low- and high-frequency ends of this range occurring much more abruptly at the low-frequency border.

Jingzhou Yang, Jia Liu [2], present Query by humming (QBH) is an efficient way to search the song from a large database. We proposed a note-based system, which consists of noted-based linear scaling (NLS) and noted-based recursive align (NRA), and makes use of both note information and the difference among the distributions of humming. Comparison experiments against several other widely-used algorithms in QBH reveal that our proposed system can get a good balance between computation time and recognition rate.

The paper [3] presents probabilistic approach to template-based chord recognition in music signals. The algorithm only takes Chroma gram data and a user-defined dictionary of chord templates as input data. Training or musical information such as key, rhythm, or chord transition models is not required. The concept of template-based chord recognition was proposed by Laurent Oudre and Cedric Fevotte[3]. The chord occurrences are treated as probabilistic events, whose probabilities are learned from the song using an expectation–maximization (EM) algorithm. The chord transcription output by our automatic chord transcribe is a sequence of chord labels with their respective start and end times. This output can be used for song playback - which constitutes the main aim of our system, but also in other applications such as song identification, query by similarity analysis.

A new method called feed-forward neural network has been proposed [4] to be used for chord recognition from input audio signal. The method uses the known feature vector for automatic chord recognition called the Pitch Class Profile (PCP). Although the PCP vector only provides music attributes corresponding to 12 semitone values, they show that it is adequate for chord recognition. The use of a simple 12-bin PCP vector based on the Discrete Fourier Transform, we show promising results and fast processing, which would have probably not been achieved with more complex pre-processing steps.

S.Suguna, J.BeckyElfreda [5], proposed cepstral Features for audio retrieval. Audio information retrieval has been performed on GTZAN datasets using weighted Mel-Frequency Cepstral Coefficients (WMFCC) feature which is a kind of cepstral feature. The results obtained for the various stages of feature extraction WMFCC and retrieval performance plot has been presented. The use of Distance-from-Boundary (DFB) and Support vector machine (SVM), audio retrieval and classification task which use Mel cepstralfeature had been performed on a database which consists of 409 sounds of 16 classes.

III. MUSICAL FUNDAMENTAL

3.1 Musical Notation

Musical notation is the representation of sound with symbols. Music can be represented using these symbols. The basic notes in music are C, D, E, F, G, A and B. A pause in music is represented by the following Figure 1 represents the musical notation from C to B.

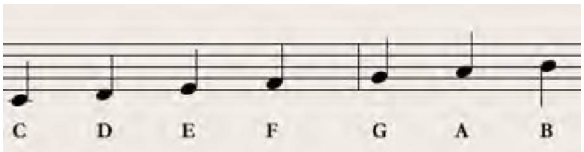

Fig.1 Musical notation from C to B

To convert the sheet music to binary digital form, consisting of only 0s and 1s, we use the following substitution.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

Table .1 Binary Conversions

Music note	Binary
C	000
D	001
E	010
F	011
G	100
A	101
B	110
-	111

3.2 KEY

In music a KEY is the major or minor scale around which a piece of music revolves. A song in a major key is based on a major scale. A song in a minor key is based on a minor scale. The Key determines how many sharps or flats are in the piece. For example if it's in the key of C major then the music will have no sharp or flat notes in it. If it was in the key of G the music would have one sharp in it is F#.

Major KEY	A	B	C	D	E	F	G
Minor KEY	Am	Bm	Cm	Dm	Em	Fm	Gm

Fig.2 Major and Minor Key

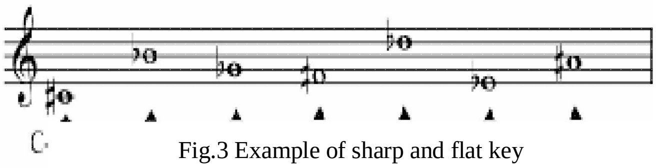

Fig.3 Example of sharp and flat key

3.3 Chords

A chord is a combination of three or more notes. Chords are built off of a single note, called the root. A chord occurs when multiple notes (Frequencies) are played simultaneously. The sequence of chords determines the chord progression. Chord recognition means transcription of a sound into a chord which can be classified into different types of chords such as major, minor, augmented, and diminished.

Table .2 Chords Type

Chord Type	Symbol	Notes
Major	C,CM, Cmaj	C E G
Minor	Cm, Cmin,Cmi	C E♭ G
Augmented	Caug, C^+, C+	C E G♯
Diminished	Cdim, C^o, Cm(♭5)	C E♭ G♭

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

IV. PROPOSED METHODOLOGY

4.1 Proposed System

The idea is to study chords - a harmony-related mid-level feature for the task of scalable CBMIR and exploit chord progressions (CPs) to realize accurate summarization of music content and efficient organization of the database. CPs means transitions between adjacent chords which contains music information related to tonality and harmony and this information is helpful for effectively distinguishing whether music audio tracks are similar to each other or not. But it is easy to extend the idea to longer chord progressions.

Steps in proposed approach are as follows:

1. Take Audio as a Input (.mp3)
2. Perform Feature Extraction on input audio signal
3. recognizing CPs from a music audio track based on a supervised learning model
4. Organizing the summaries of audio tracks in the database using an indexing structure and it's Chord Progression Histogram (CPH) is computed
5. Perform similarity searching between dominant chord progressions (CPs)
6. If match
7. Retrieval results (relevant songs are returned to the user)
8. Else
9. Search in another file in database.

In Figure 4, shows flow of Proposed Approach

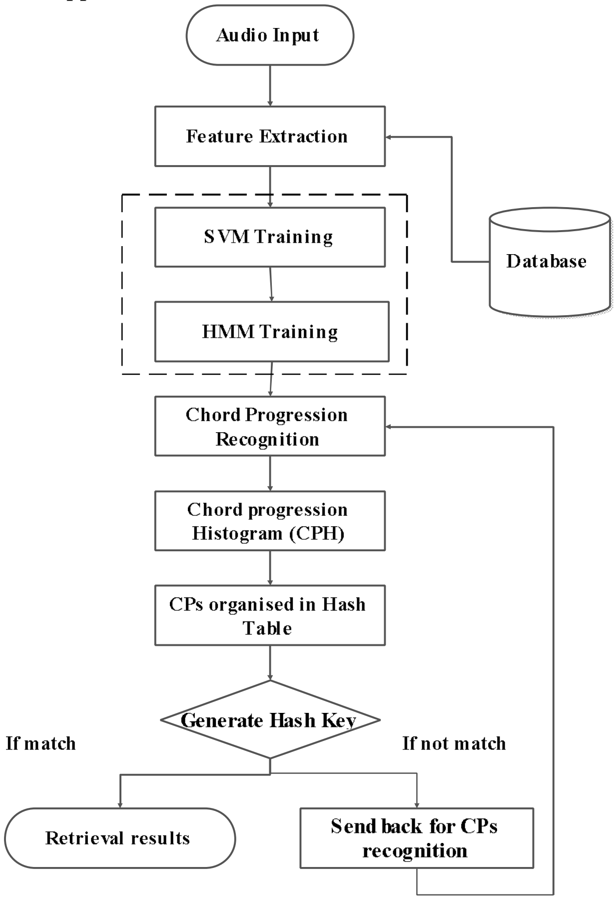

Fig.4 Proposed System Flowchart

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

4.2 Audio Retrieval System Architecture

The basic operation of the system is as follows. First, the feature vectors are estimated for each audio track from the database. Then, for each input audio track from the user, its feature vectors are estimated. Then supervised learning model used for recognizing CPs from each Feature vector from input audio track. Similarly, CP recognition is performed for all songs in the database. Their Chord Progressions Histogram (CPHs) are computed and organized in the hash table, where the set of dominant CPs of each song is used as its hash key. With a query as input, its CPs is recognized and its CPH is computed. With its hash key, relevant songs are found from the associated buckets. Finally, based on similarity between CPs of input audio track and CPs of all songs in the database relevant songs are returned to the user in a ranked list as the retrieval results.

Figure 5 shows System Architecture for audio retrieval system. It consists of four main parts: chord model training, CP recognition from audio, CPH computation and hash table organization.

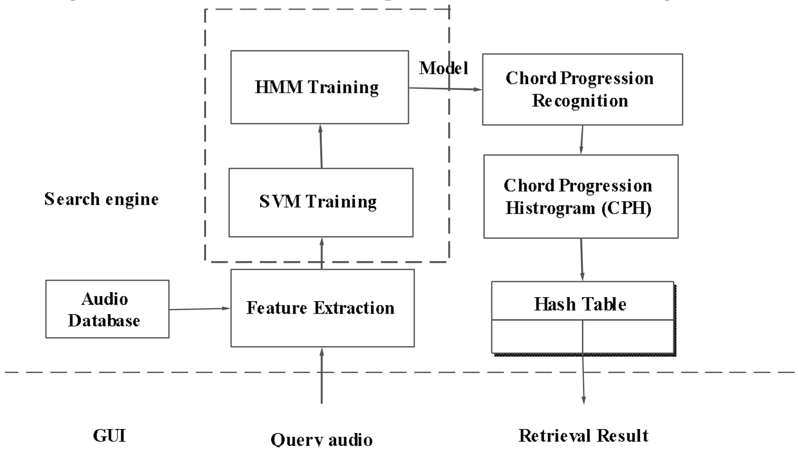

Fig.5 System Architecture for audio retrieval

4.3 Feature Extraction

Feature Extraction is the process of computing a compact numerical representation that can be used to characterize a segment of audio. Feature extraction involves the analysis of the input of the audio signal. Chroma has been the most successfully used feature for the chord recognition task. It consists of a sequence of chroma vectors. Each chroma vector, also called Harmonic Pitch Class Profile (HPCP) vector, describes harmonic content of a given frame. Since a chord consists of a number of tones and can be uniquely determined by their positions, HPCP vector can be effectively used for the chord representation. The most common way of calculating chroma vector is to transform the signal from the time domain to the frequency domain with the help of Fast Fourier transform (FFT).Figure 6, shows steps in HPCP feature extraction process which is summarized as follows:

1. First start by cutting the song into short overlapping and windowed frames.
2. Perform a spectral analysis (To know the frequency components of the music signal) using the discrete Fourier transform (DFT) to convert the signal into a spectrogram.
3. Compute a set of local maxima or peaks and select the frequency values between 100 and 5000 Hz using frequency filtering.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

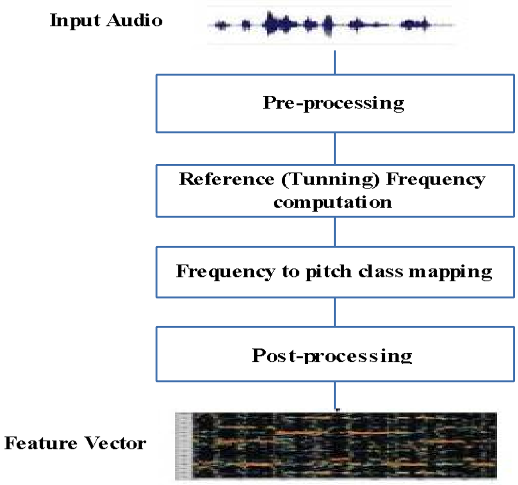

Fig. 6 HPCP feature extraction process

4. Perform reference frequency computation procedure.
5. Do pitch class mapping with respect to the estimated reference frequency. This procedure is used for determining the pitch class value from frequency values. A weighting scheme with cosine function is used and it considers the presence of harmonic frequencies.
6. Then do normalization. (To normalize the each feature frame by frame dividing through the maximum value to eliminate dependency on global loudness).
7. Get a result HPCP vector.

4.4 Chord Progression (CPs) Recognition Method

A chord is a set of harmonically related musical pitches that sounded almost simultaneously, and the sequence of chords determines the chord progression. Chord recognition means transcription of a sound into a chord which can be classified into different types of chords such as major, minor, augmented, and diminished. Recognizing CPs from a music audio track based on a supervised learning model such asSupport vector machine (SVM) which is used for chord detection and Hidden Markov Models (HMM) which is used for chord progression. Multi-probing is used in chord progression recognition via the modified Viterbi algorithm, which outputs multiple likely chord progressions and increases the probability of finding the correct one. A chord progression histogram (CPH) is computed from each music audio track and stored in hash table.

4.5Locality Sensitive Hashing

Locality Sensitive Hashing (LSH) is an index-based data organization structure which is used to quickly and approximately find items relevant to a given input query. Here only one LSH table with tree structure is using for Organization of audio tracks according to their CPHs and a set of dominant chord progressions (CPs) of each song is used as the hash key.

4.6 Similarity Searching

Each song has its own CPs. Two similar songs share many CPs in common. Therefore, it is possible to use CPs in the hash design. Longest common chord subsequence (LCCS) is used to measure the similarity between two chord sequences.The chord sequence recognized from the feature sequence is a mid-level representation of an audio signal. Directly comparing two chord sequences is faster than comparing two chroma sequences. But it still requires time-consuming. To process the retrieval process with amore compact representation, the chord sequence can be further summarized into a chord progression histogram.Estrada distance [8] uses textures to compare two chords, and then to compute a musical distance to measure transitions between chords. A texture is obtained from a chord by ordering the different pitches on the same octave, adding the lowest pitch one octave higher. A distance between two chords is computed following the formula:

$$\delta(C1,C2) = \max(\#(C1),\#(C2)) - |VC1 \cap VC2| - 1 \qquad (1)$$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

Where C1 and C2 are two chords, #(C1) and #(C2) their respective number of notes and VC1 and VC2 their respective interval vectors. Negative results are set to 0. The distance values come from 0 to 11.
In Figure 7 shows feature called chord progression histogram (CPH) to show the percentage of time each CPs occupies in a song.

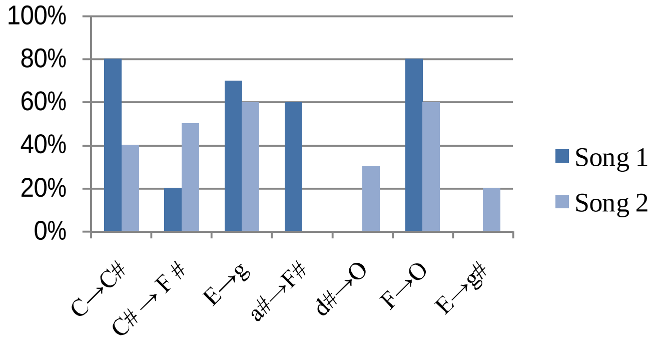

Fig.7 A chord Progression histogram sample showing the CPs similarity between songs two songs (for demonstration purpose only)

In Fig 7, we can see that the CPs C→C#,C#→ F#, E →g, F→O and Am frequently appear in both songs.
In Figure 8, shows similarity between two songs.

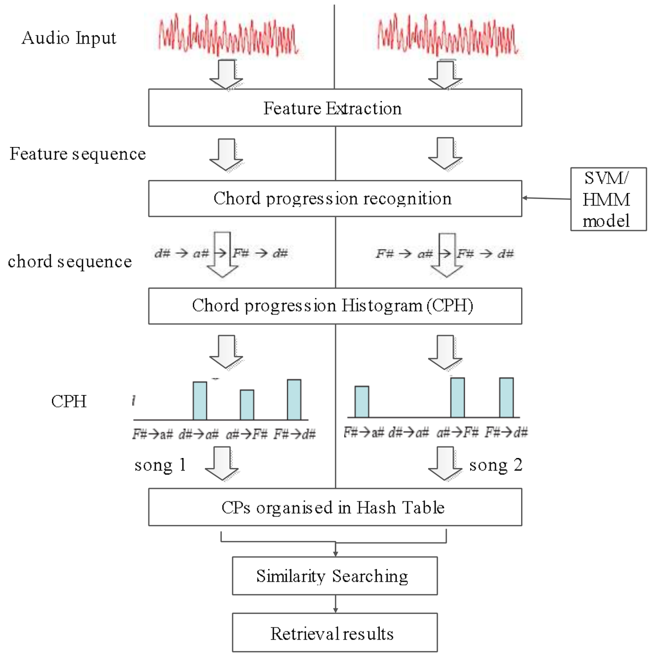

Fig.8 Process for comparing two songs

V. CONCLUSION

The use of content-based approaches has always been a common strategy in the music information retrieval area. Content Based Audio Retrieval (CBAR) is one of the most popular research areas on social web sites. The algorithm consists of two key points: recognizing chord progressions from a music audio track based on a supervised learning

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 4, Issue 1, January 2016

model related to musical knowledge and computing a summary of the audio track from the recognized chord progressions by locality sensitive hashing (LSH). Statistical approaches for extraction of chord progressions using SVM and HMM based framework is proposed. Multi-probing is used in chord progression recognition via the modified Viterbi algorithm, which outputs multiple likely chord progressions and increases the probability of finding the correct one. A chord progression histogram is used to summarize the probed CPs in a concise form, which is both efficient and also retains local chord progressions. Precision, Recall and accuracy are used for performance measure.

REFERENCES

[1] Heng-Tze Cheng, I-Bin Liao and Homer H. Chen, “Automatic Chord recognition for music classification and retrieval”, ICME 2008, pp. 1505-1508.
[2] Jingzhou Yang, Jia Liu, Wei-QiangZhang,“A Fast Query by Humming System Based on Notes”, In Proc. ISCA, September 2010, pp. 2898-2901.
[3] Laurent Oudre, Cedric Fevotte, “Probabilistic Template-Based Chord Recognition”, IEEE November 2011, pp. 2249 – 2259.
[4]J. Osmalskyj, J.J. Embrechts, S. Pierard and M. Van Droogenbroeck, “Neural Networks for Musical Chords Recognition”,JIM May 2012, pp. 9-11.
[5] S.Sugunaand J.BeckyElfreda, “Audio Retrieval based on Cepstral Feature”, In Proc. IJCA,December 2014, pp. 28-33.
[6] J. T. Foote, “Content-Based Retrieval of Music and Audio”, In Proc. SPIE, vol.3229, 1997, pp. 138-147.
[7] Wang. A, “Automatic Identification of Musical Versions Using Harmonic Pitch Class Profiles”, PhD thesis, September 2011.
[8] MaksimKhadkevich, “Music signal processing for automatic extraction of harmonic and rhythmic information”, PhD thesis, December 2011,
[9] T. E. Ahonen, “Combing chroma features for cover version identification,” In ISMIR,2010, pp. 165–170.
[10] Y. Yu, M. Crucianu, V. Oria, and E. Damiani, “Combing multi-probing histogram and order-statistics based LSH for scalable audio content retrieval,” In Proc. ACM MM, October 2010, pp. 381–390.
[11]Y. Yu, R. Zimmermann, Y.Wang, and V. Oria, “Recognition and summarizationof chord progressions and their application to music information retrieval,” in Proc. IEEE ISM, 2012, pp. 9–16.
[12] H.T.Cheng, Y.-H.Yang, Y.-C. Lin, I.-B.Liao, and H.H. Chen, “Automatic chord recognition for music classification and retrieval,” in Proc.ICME, 2008, pp. 1505–1508.