A Music Information Retrieval Algorithm using MP and SPM

Soe Myat Thu
University of Computer Studies,Yangon
thuthu052228@gmail.com

Abstract

In the IT age where private music collections contain thousands of songs and commercial music catalogs consist of millions of songs, music recommender systems become more and more important especially for searching and browsing music catalogs. In this paper, retrieving the required information from acoustic music signal in an efficient way is considered. The approach is to segment those audio signals and determined similarities among songs, particularly, a piece of input music signal compared with storage music song's signal into the database and then to retrieve the similar song. Representing the audio signal having sparse nature is accomplished by Matching Pursuit. In order to matching a candidate segment with the query segment, the audio signal similarity measure is performed by Spatial Pyramid Matching.

1. Introduction

A significant amount of music information retrieval (MIR) research has focused on signal processing techniques that extract features from audio content. Often, these feature sets are designed to reflect different aspects of music such as timbre, harmony, melody and rhythm.In addition, researchers use data from online sources that places music in a social context. Individual sets of audio content and social context features have been shown to be useful for various MIR tasks (e.g., classification, similarity, recommendation). Among them, Similarity is crucial for the effectiveness of searching music information and the music segmentation. There exist three general recommendation approaches, namely the collaborative filtering approach, the content-based approach and hybrid approaches.

In the early years of MIR (1985-95), research concentrated on rudimentary time and frequency domain features such as (1) windowed amplitude data and derived tempo statistics, and (2) windowed spectra, reduced spectra, and derived spectral statistics ("spectral measures"). The next generation of MIR (roughly 1995-2005), saw the introduction of more sophisticated features involving higher level time and frequency domain features such as beat histograms, Mel-frequency cepstral coefficients (MFCCs) and chromagrams. In addition, researchers began using more sophisticated statistics to aggregate the values of each feature within a song, and using newer machine learning techniques.

Automatic analysis of the structure has been studied mainly for the application of creating a meaningful summary of a musical piece. One of the first works operating on acoustic signals was by Logan and Chu, describing an agglomerative clustering and hidden Markov model (HMM) based approaches for key phrase generation [1]. They used mel-frequency cepstral coefficients (MFCCs) from short (26 ms),overlapping frames.The clustering method grouped the frames together iteratively until a level of stability had been reached. In the HMM method they trained an ergodic HMM with only few states, hoping that each state would represent a musical part, and used the Viterbi decoded state sequence as the description of the musical structure. The HMM approach was taken further by Aucouturier and Sandler using spectral envelope as the feature [2]. It was noted in both these studies that when using such short frames, the HMM states did not model musically meaningful parts, as was hoped. Abdallah et al increased the frame length considerably and the number of states up to 80 [3]. After acquiring the state sequence, each frame was provided with some knowledge about the surrounding context by calculating a state histogram in a 15 frame window. The histograms were then used in clustering the frames by optimising a cost function with simulated annealing. Rhodes et al added a term to control the duration of stay in a certain cluster [4], while Levy et al refined the clustering method to a context aware variant of fuzzy C-means [5]. Another popular starting point of the analysis is to calculate frame-by-frame similarities over the whole signal, constructing a self-similarity matrix. Foote proposed to use the similarity matrix for visualising music [6]. It was noted that the parts of music having similar timbral characteristics created visible areas in the similarity matrix. The borders of these areas were sought and used in segmenting the piece in [7]. In [8] Foote and Cooper used a spectral clustering method to group similar

segments.When the used feature describes the tonal (pitch) content of the signal instead of general timbre, e.g., chroma instead of MFCCs, repetitions generate off-diagonal stripes to the similarity matrix instead of rectangular areas of high similarity. Such stripes reveal similar sequential structures, e.g.melody lines or chord progressions, instead of just denoting parts having similar timbral characteristics, or sounding the same. The two main approaches (HMM-based "state"method and "sequence" method relying on stripes in the similarity matrix) were compared by Peeters [9]. He noted that as the sequence approach requires a part to occur at least twice to be found, the HMM approach would be more robust analysis method. Still, the stripes have been used in structure analysis by several authors. Bartsch and Wakefield extracted chroma from beat-synchronised frames and used the most prominent off-diagonal stripe to define a thumbnail for the piece [10]. Lu et al proposed a distance metric considering the harmonic content of sounds, and used 2D morphological operations (erosion and dilation) to enhance the stripes [11]. In popular music pieces, the clearest repeated part is often the chorus section. Goto aimed at detecting it using chroma, and presented a method for handling the musical key modulation sometimes taking place in the last in the last refrain of the piece [12]. Music tends to show repetition and similarities on different levels, starting from consecutive bars to larger parts like chorus and verse. Some authors have tried to take this into account and proposed methods operating on several temporal levels. Jehan constructed several hierarchically related similarity matrices [13]. Shiu et al extracted chroma from beat-synchronised frames and then used dynamic time warping (DTW) to calculate a similarity matrix between all the measures of the piece [14]. The higher level musical structure was then modelled with a manually parametrised HMM. Dannenberg and Hu gathered the shorter repeated parts and gradually combined them to create longer, more meaningful, parts in [15]. Later, Dannenberg used the stripes in similarity matrix to find similar musical sections, and then utilised this information to aid a beat tracker [16]. Chai proposed to take the context into account by matching two windows of frame level features with DTW. Sliding the other window while keeping the other fixed provided a method to calculate the similarity on different lags and to determine the lag of maximum similarity. Gathering this information in a matrix formed stripes of prominent lags, like the stripes in a similarity matrix. The longer stripes were then interpreted information about the repeats of structural parts [17].Maddage et al proposed a method for analysing a musical piece combining different sources of information. They used beat-synchronised pitch class profile as the feature and detected chords with pre-trained HMMs. Using assumptions of the lengths of the repeated parts, fixed length segments were matched to get a measure of similarity. Finally heuristic rules, claimed to apply on English-language pop songs,were used to deduce the high-level structure of the piece [18].

This paper presents a content-based approach to determine similar song from a database and retrieve a whole music song according to the input query. Because of the challenge of matching a candidate segment with the query segment, the system could significantly improve similarity measure using Spatial Pyramid Matching. And the retrieval time could considerably improve using Matching Pursuit Method. Our particular approach to choose music song also makes it possible automatically retrieval using matching pursuit features sets, for example for use in browsing rapidly through a list of possible song of interest returned by a search engine. By guiding us to the most significant parts of a music song, it also allows the development of fast and efficient methods for searching very large collections based purely on the audio content of the song, sidestepping the computational complexity of existing content-based search methods.

2. Background

2.1 Matching Pursuit

Matching Pursuit is part of a class of signal analysis algorithms known as Atomic Decompositions. These algorithms consider a signal as a linear combination of known elementary pieces of signal, called atoms, chosen within a dictionary.

In the framework of music signal analysis, it is desirable to obtain sparse representations that are able to reflect the signal structures, eg. issued from musical songs. MP aims at finding sparse decompositions of signals over redundant bases of elementary waveforms. Wavelet transforms should be designed as follow: Dictionary: A dictionary contains a collection of blocks plus the signal on which they operate. It can search across all the blocks (i.e., all the scales and all the bases) for the atom which brings the most energy to the analyzed signal. Book: A book is a collection of atoms. Summing all the atoms in a book gives a signal. Our implementation of the Matching Pursuit algorithm uses roughly 3 steps,

1. Update the correlations in the blocks, by applying the relevant correlation computation algorithm to the

analyzed signal, and find the maximum correlation in the same loop.
2. Create the atom which corresponds to the maximum correlation with the signal (and store this atom in the book).
3. Subtract the created atom from the analyzed signal, thus obtaining a residual signal,and re-iterate the analysis on this residual[MP].

Using Matching Pursuit method is price of efficiency and convergence. Time compression is quite excellent by extracting prominent atoms (features). MP is also used visualization and transformating audio signal [19], harmonic decomposition of sounds [20]. In order to achieve the required information in our system, the algorithm could use as follow steps:

1. initialization:
2. computation of the correlations between the signal and
every atom , using inner products :
3. search of the most correlated atom, by searching for the maximum inner product:
4. subtraction of the corresponding weighted atom from the signal :
5. If the desired level of accuracy is reached, in terms of the number of extracted atoms or in terms of the energy ratio between the original signal and the current residual, stop; otherwise, re-iterate the pursuit over the residual: and go to step 2.

2.2. Spatial Pyramid Matching

Spatial Pyramid Matching is to find an approximate correspondence between these two sets by level. At each level of resolution, it works by placing a sequence of increasingly coarser grids over the features.

A pyramid match kernel allows for multiresolution matching of two collections of features in a high-dimensional appearance space, but discards all spatial information. Another problem with this approach is that the quality of the approximation to the optimal partial match provided by the pyramid kernel degrades linearly with the dimension of the feature space. To overcome these shortcomings, we propose instead to perform pyramid matching in the feature space.

In this approximate matching approach, SPM is constructed the pyramid level and then the number of matches at level L is given by histogram intersection function. In Similarity, SPM is used step by step level to improve matching musical data space and taking a weighted sum of the number of matches. At any fixed resolution, two points are said to match if they fall into the same cell of the grid. The number of matches at each level is given by the *histogram intersection* function.

Using spatial pyramid matching is its efficiency, its use of implicit correspondences that respect the joint statistics of co-occuring features, and its resistance to 'superfluous' data points. Convergence is also guaranteed since pyramid match kernel is positive-definite function. Model free and effective for finding sparse over-complete representation.

3. Proposed System

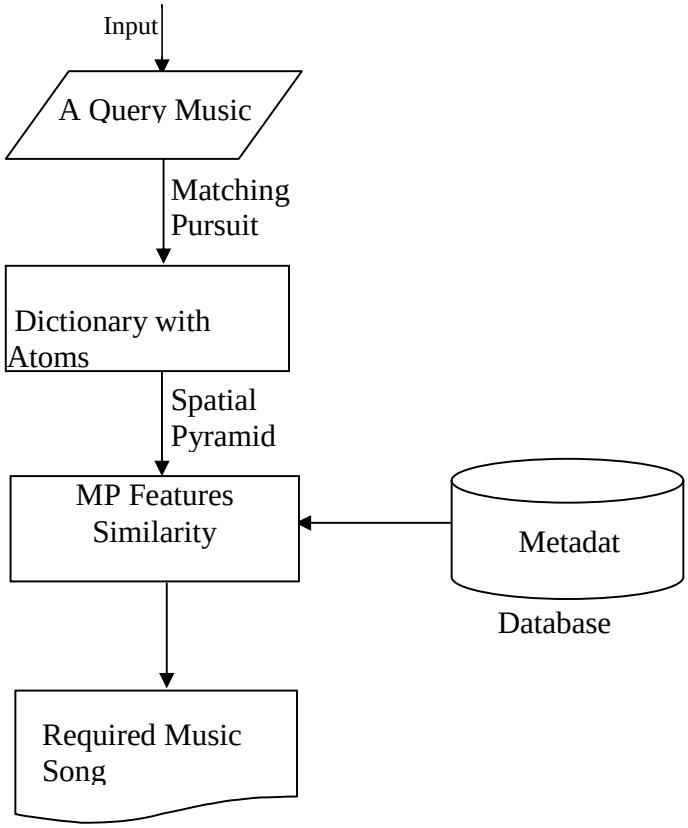

Figure1: A block diagram of the system.

A block diagram of the proposed system can be seen in Figure1.The proposed method creates matching pursuit features from a query input music signal. Music signal structural features are represented dictionary with most prominent atoms that match their time-frequency signature. And finds the optimal description of the music parts from the meta database in respect to the faster and more similar function defined in Spatial Pyramid Matching method as described in Section 2. Using Matching

Pursuit and Spatial Pyramid Matching method in this proposed music information retrieval method, the search can be optimised by occupying different groups in order which eliminates much of the search space.

4. Evaluation Study

Music songs are meaningful and no societies without music. In religion, sports and work, music songs are essential for social activities .In music information retrieval system, new search systems study and analyze the problem of music browsing to become more and more similarity and effectively. Using MP and SPM method, this approach would be more effective and efficient than existing methods in retrieving similar music information. The performance of the system will be evaluated in simulations browsing the similar structure of a set popular music pieces.

5. Conclusion

In music information retrieval, browsing and search particular music songs in an efficient manner is still demanding. This paper proposed a framework for new search engine model in music information retrieval. The approach system would use matching pursuit and spatial pyramid matching for determining significant features of music pieces and retrieving music queries in efficient way. The feature sets will be achieved by matching pursuit method as training and testing data. Retrieving similar music pieces from a database is completed by matching the feature space by step by step level using spatial pyramid matching. Better speed and accuracy can be expected upon the whole architecture of the system.

6. References

[1]. B. Logan and S. Chu." Music summarization using keyphrases*". In Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, pages749–752, Istanbul, Turkey, June 2000.*

[2]. J,J.Aucouturier and M. Sandler. "Segmentation of musical signals using hidden Markov models*". In Proc. of 110th Audio Engineering Society onvention, Amsterdam, The Netherlands, May 2001*.

[3]. S. Abdallah, K.Nolad, M. Sandler, M.Casey, and Rhodes." Theory and evaluation of a Bayesian music structure extractor.*" In Proc. of 6th International Conference on Music Information Retrieval, London, UK, Sept. 2005*.

[4]. C. Rhodes, M. Casey, S. Abdallah, and M. Sandler." A Markov-chain Monte-Carlo approach to musical audio segmentation." *In Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, pages 797–800, Toulouse, France, May 2006.*

[5]. M. Levy, M. Sandler, and M. Casey." Extraction of high-level musical structure from audio data and its application to thumbnail generation." *In Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, pages 13–16, Toulouse, France, May 2006.*

[6]. J. Foote." Visualizing music and audio using self-similarity." *In Proc. of ACM Multimedia, pages 77–80, Orlando, Florida, USA, 1999.*

[7]. J. Foote. "Automatic audio segmentation using a measure of audio novelty." *In Proc. of IEEE International Conference on Multimedia and Expo, pages 452–455, New York, USA, Aug. 2000.*

[8]. J. T. Foote and M. L. Cooper. "Media segmentation using self-similarity decomposition. "*In Proc. of The SPIE Storage and Retrieval for Multimedia Databases, volume 5021, pages 167–175, San Jose, California, USA, Jan. 2003.*

[9]. G. Peeters. Deriving musical structure from signal analysis for music audio summary generation:"sequence" and "state" approach. In Lecture Notes inComputer Science, volume 2771, pages 143–166.Springer-Verlag, 2004.

[10]. M. A.Bartsch and G.H Wakefield. To catch a chorus: "Using chroma-based representations for audio thumbnailing." In Proc. of 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pages 15–18, New Platz, New York, USA, Oct. 2003.

[11]. L. Lu, M. Wang, and H.-J. Zhang. "Repeating pattern discovery and structure analysis from acoustic music data." *In Proc. of Workshop on Multimedia Information Retrieval, pages 275–282, New York, USA, Oct. 2004.*

[12]. M. Goto." A chorus-section detecting method for musical audio signals." *In Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing, pages 437–440, Hong Kong, 2003*.

[13]. T. Jehan." Hierarchical multi-class self similarities." *In Proc. of 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pages 311–314, New Platz, New York, USA, Oct. 2005.*

[14]. Y. Shiu, H. Jeong, and C.-C. J. Kuo. "Musical structure analysis using similarity matrix and dynamic programming." *In Proc. of SPIE Vol. 6015 - Multimedia Systems and Applications VIII, 2005.*

[15]. R. B. Dannenberg and N. Hu. "Pattern discovery techniques for music audio. "*In Proc. of 3rd International*

Conference on Music Information Retrieval, pages 63–70, Paris, France, Oct. 2002.

[16]. R. B. Dannenberg. "Toward automated holistic beat tracking, music analysis, and understanding." *In Proc. of 6th International Conference on Music Information Retrieval, pages 366–373, London, UK, Sept. 2005.*

[17]. W. Chai. "Semantic segmentation and summarization of music: methods based on tonality and recurrent structure." IE*EE Signal Processing Magazine, 23(2):124–132, Mar. 2006.*

[18]. N. C. Maddage, C. Xu, M. S. Kankanhalli, and X. Shao. "Content-based music structure analysis with applications to music semantics understanding*." In Proc. of ACM Multimedia, pages 112–119, New York, New York, USA, Oct. 2004.*

[19]. Garry Kling and Curtis Roads, "Audio analysis, Visualization, And Transfromation with The Matching Pursuit Algorithm", *proc. of the 7th Int.Conference on Digital Audio Effects,Naples,Italy,October5-8,2004.*

[20]. Sacha Krstulovic,Remi Gribonval, "A Comparison of Two Extensions of the Matching Pursuit Algorithm for The Harmonic Decomposition of Sounds", *2005 IEEE Workshop Application of Signal Processing to Audio and Acoustics.*