A Survey on Audio Retrieval System for Classification

Priyanka S. Jadhav., Saurabh H. Deshmukh

Department of Computer Engineering, G.H.Raisoni College of Engineering and Management, wagholi, Pune, India

ABSTRACT: In today's environment, most of the retrieval algorithms are textual based algorithm therefore we cannot able to make classification of musical instruments. In most of the retrieval system the classification can be done on the basis of term frequencies and use of snippets in any documents. Existing search engines (like Yahoo, Google, AltaVista etc.) make similarity search on the basis of Key-word and snippets, but sometimes user may not able to express the queries in words so we have to switch to audio retrieval system. In existing audio visual retrieval system,Content-Based retrieval systems user can enter any queries ranging from drawing sketch to sing a song or a video clip or set of images from some video short for a video retrieval. Which add more comprehensive approach for users to enter their queries .From [11]we can say that content-based retrieval system permits more tolerance towards erroneous queries, as in these systems queries contain more errors; so for such search keys similarity search based on approximate matching produce batter results compare to exact matching. In existingsystem, systems are classified based on audio object representation, indexing structure and retrieval technique used. In our proposed system we are extracting the sound by recognizing the timbre of sound. In many existing audio retrieval system we extract the features either by linear predictive code or by perceptual linear prediction. But in proposed system for extraction we use musical information retrieval toolbox (MIR toolbox) which is useful to find out audio descriptor by using hybrid selection method. After finding audio descriptor we identify the musical instruments with the help of vector quantization

KEYWORDS: Term frequencies, content based image retrieval system, MIR toolbox etc.

I.   INTRODUCTION TO EXISTING SYSTEMS

In existing system, for any visual content retrieval we have to extract both the audio & video features. The extraction could be done by the method such as

A.   Temporal Features:
The temporal domainis the natural domain for audio signals. All temporal features have in common that they are mined directly from the raw audio signal, without any previoustransformation. Consequently, the computational complexity of temporal features tends to be short. To extract the temporal features [11] we partition the group of temporal features into three groups, depending on what the feature describes
1.   zero crossings feature
2.   Amplitude Base features
3.   power base features
In zero crossing features, Zero crossings are the basic property of an audio signal that is often employed in audio classification. Zero crossingsallow for a rough estimation of dominant frequency and the spectral centroid it's having three different phases
➢   Zero Crossing Rate
➢   Linear Prediction Zero crossing Rate
➢   Zero crossing peak amplitude
Where as in amplitude base features, many features are directly computed from the amplitude i.e. the pressure variation of a signal. Amplitude-based featuresare easy and fast to compute but limited in their articulateness. They represent the temporal envelope of the audio signal. It has two phases
         1: MPEG-7 audio waveform (AW)
         2: Amplitude descriptor (AD).

And in power based features, the energy of a signal is calculated as the square of the amplitude represented by the waveform. The power of a sound is the energy transmitted per unit time. Consequently, power is the mean-square of a signal. Many times the root of power (root-mean-square) is used for feature extraction. In case of existing textual search engine, a web crawleris maintained to search the document. The queried keyword is searched against database documents by measuring the semantic similarities parameters such as removing the stop words i.e. preprocessing of data to reduce the complexity of system. Finding total no. of snippet count in the given document. And finding the term frequency to rank the data. Where as in existing musical retrieval system we use

1. **Signal Parameter Based Modeling:**
This type of modeling is applicable to musical objects as well as audio objects, because both can be characterized by signal or acoustical parameters[11]. To model an object we need both frame level as well as global parameters. General attributes required for this type of modeling are, zero-crossing rate, energy, pitch frequency, timber, energy contour, and loudness contour etc.

2. **Vector Space Bases Retrieval or Vector Based Model**
Both the query and each[10] object are characterized as vectors in terms of n-dimensional space. A measure of similarity between the query and each object in the database is computed.

3. **Pattern Matching Based Retrieval or String Matching Bases Retrieval**
In pattern matching based retrieval [10], both the queries and the document are represented by a sequence of characters, integers, words etc., and similarity between them is computed based on, how similar two sequences are? Similarity matching between them can be determined either by exact sequence matching or by approximate sequence matching.

4. **Sequence-Comparison using Dynamic Programming (DP):**
There are number of methods for sequence comparison, but sequence matching using DP is quite popular than all other due to its space efficient implementation and with lower complexity. Sequence matching based on DP uses the concept of edit distance; edit distance is the cost of changing the source sequence (source string) into target sequence (target string). The cost of transposing source string into target string is calculated in terms of edit operators, common edit operators used in DP are substitution (replacement), deletion and insertion.

II.   LITERATURE SURVEY

Audio parameters based systems have been extensively used for speak recognition and speaker identification systems for more than two decades and these systems are still popular in this area, but unfortunately, for CBAIR systems based especially for music retrieval systems; audio parameter based systems could not gain same popularity. Main reason might be, that these systems did not support QBH, which is very popular in the area of music retrieval now a day. During past few years many researchers developed systems for CBAIR based on audio parameters that support QHE only.

J. T Foote in [12] presented an idea for the representation of an audio object by a template that characterizes the object in his purposed system. For construction of a template; an audio signal is first divided into overlapping frames of constant length then using simple signal processing techniques, for each frame a13-dimensional feature vector is extracted (12 Mel-Frequency Cepstral Coefficients plus Energy) at a 500Hz, and then these feature vectors are used to generate templates using tree-based Vector Quantized trained to maximize mutual information (MMI). For retrieval, query is first converted in to template in the same way described earlier then for its similarity search template matching is applied which uses distance measure, and finally a ranked list is generated based on minimum distance. In this system performance of the system with Euclidean distance as well as Cosine distance, is also compared, and experimental results show that cosine distance performs slightly better than Euclidean distance. This system may fail for music retrieval if either query is corrupted with noise or bad quality recorded.

Muscle fish group [13], in this system an audio object is characterized by its frame level and global acoustical and perceptual parameters. These features are extracted at frame level using signal processing techniques and globally using statistical analysis based on frame level features and musical features (for music signals only) using musical analysis. Frame level features consist of loudness, pitch, tone (brightness and bandwidth), MFCCs and derivative. Global

features are determined by applying statistical modeling techniques on the frame level features that is, using Gaussian and Histogram Modeling techniques to analyze audio objects. For musical objects, musical features (i.e. rhythm, events and distance (interval)) are extracted using simple signal processing techniques like pitch tracking, voiced and unvoiced segmentation and note formation. For indexing, multidimensional features space is used. For retrieval, distance measure is used and to improve the performance, a modified version of query-point-expansion technique is used, but here expansion for the refinement of the concept if achieved by standard deviation and mean of the objects in the expected region. This system again bounded by its inherited limitation, and works for QBE only.

G. Li and A. Khokhar [14] represented an audio object by 3-dimensional feature vector for their system. Feature vectors are extracted at frame level using Discrete Wavelet Transform (DWT). They applied 4-level DWT decomposition to audio signal, then in transformed domain variance, zero-crossing rate and mean of wavelet coefficients are determined to form feature vectors. For indexing structure, B-tree structure is used, which is constructed using clustering technique along with multiresolution property of the wavelet transform. Similarity search is applied using weighted Euclidean distance, and based on minimum distance a ranked target list is retrieved for the desired query.

S. R. Subramenya and A. Youssef [15] presented a signal processing based approach using Discrete Wavelet Transform (DWT) for feature vector of an audio object. First of all an audio signals is decomposed using 4-level DWT and then form wavelet coefficients, feature vector is formed using all approximate coefficients and 20%-30% of detail coefficients of all levels obtained during wavelet decomposition. For query processing same procedure is applied. They did not specified indexing technique but for similarity matching they used Euclidean distance measure.

AsifGais et .al.[16], their system is considered to be the first complete CBMIR database system based on QBH, it was consist of 183 songs in MIDI file format. To represent music object, songs in MIDI format are first converted into melodic contours and then each contour is transformed into a string of characters U(up), D(down) and S(same). To generate query, pitch is extracted from recorded hummed query then it is convert into melodic contour which is converted into a string of same three characters (U, D, and S). For retrieval, similarity for input query is evaluated using approximate string matching algorithm then a rank list of similar melodies(songs) based on the minimum edit distance is generated. Performance of this system was quite satisfactory, and one of the reasons for good performance is its very small size music database.

III.   PROPOSED SYSTEM

In our proposed system we are going to extract the features of sound by recognizing the timbre of sound. After feature extraction we are making classification of sound on the basis of extracted features. For retrieving the audio data we use MIR tool box. MIR tool box has the set of multiple functions written in matlab. Those functions are used to extract the audio related features. In our proposed system our main aim is to find out the audio descriptor. To find out an audio descriptor from given data we use extracted features and hybrid selection method. In hybrid selection method, we select the correct audio descriptors for the identification of singer of North Indian Classical Music. Initially only robust (primary) audio descriptors are released on the system in first pass and its impact is noted. Then only selecting the top few audio descriptors, having largest impact on the identification process, are selected and remaining will beremoved in the backward or second pass. Then choosing and freeing all the less noteworthy audio descriptors from the groups that had highest impact on singer identification process improves the chances of success of correctly identifying the singer. The method reduces substantially the large number of audio descriptors to few, important audio descriptors. The selected audio descriptors are then fed as input to further classifiers. After finding the correct audio descriptors we generate the feature vectors and we will identify the musical instruments by using vector quantization method. In vector quantization method feature vector stores the extracted features of an audio descriptor and those extracted features will be matched with another feature vector for comparison.
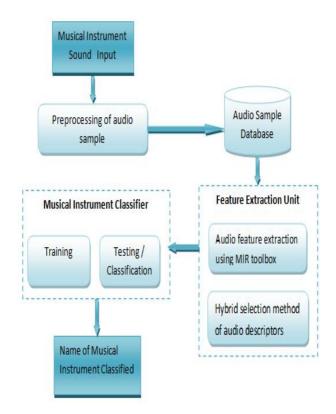
Figure 1. Block schematic of proposed system

Vector quantization is based on the competitive learning paradigm, so it is closely related to the self-organizing map model.

IV.   CONCLUSION

From the above discussion we can say that most of the retrieval systems are textual based retrieval systems and the existing audio retrieval systems are working on the principle of Signal Parameter Based Modeling, Vector Space Bases Retrieval or Vector Based Model Pattern Matching Based Retrieval or String Matching Bases Retrieval or temporal feature extractions but our system is using vector quantization method for classification of audio descriptor and audio descriptor is selected on the basis of hybrid selection model.

REFERENCES

1. D. Giannoulis and A. Klapuri, "Musical Instrument Recognition in Polyphonic Audio Using Missing Feature Approach," IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, Vols. VOL. 21, NO. 9,, SEPTEMBER SEPTEMBER 2013.
2. S. Essid and G. Richard, "Musical Instrument Recognition by Pairwise Classification Strategies," IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, Vols. VOL. 14, NO. 4, JULY 2006.
3. T. H. Park, "Towards Automatic Musical Instrument Timbre Recognition," Princeton University, New Jersey,USA, November 2004.
4. O. Lartillot and P. Toiviainen, "A Matlab Toolbox for Musical Feature Extraction from Audio," in International Conference on Digital Audio Effects, Bordeaux, France, 2007.
5. S. H. Deshmukh and S. G. Bhirud, "North Indian Classical Music's Singer Identification by Timbre Recognition using MIR Toolbox," International Journal of Computer Applications, vol. 91 – No.4, p. 0975 – 8887, 2014..
6. S. Deshmukh and S. Bhirud, "A Hybrid Selection Method of Audio Descriptors for Singer Identification in North Indian Classical Music," in Emerging Trends in Engineering and Technology (ICETET), Himji, Japan, 2012.
7. K. Teknomo, "K-Means Clustering Tutorial," July 2007. [Online]. Available: http://people.revoledu.com/kardi/tutorial/kMean/index.html
8. A. A. Wieczorkowska, Z. W. Ra´s and R. Lewis, "Multi-way Hierarchic Classification of Musical Instrument Sounds".

9.   T. Zhang, "Instrument Classification in polyphonic music based on timbre analysis".
10.   S. Essid, G. Richard and. B. David, "Musical Instrument Recognition by Pairwise Classification Strategies," IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, Vols. VOL. 14, NO. 4, JULY 2006. Shruti Vaidya1, Dr. Kamal Shah2 "Audio Denoising, Recognition and Retrieval by Using Feature Vectors" IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727Volume 16, Issue 2, Ver. II (Mar-Apr. 2014), PP 107- 112 www.iosrjournals.org
11.   J. T. Foote., "Content-Based Retrieval of Music and Audio." In Proc. SPIE, vol3229, pp. 138-147, 1997.
12.   J. Foote., "An overview of audio information retrieval." In Multimedia Systems 7, pp 2- 10. ACM, January 1999.
13.   A. Khokhar, G. Li "Content-based Indexing and Retrieval of Audio Data using Wavelet" ICME 2000
14.   S. R. Subramanya and A. Youseef, "Wavelet-based Indexing of Audio Data in Audio/ Multimedia Databases" 4th IEEE Int'l.Workshop on Multimedia DBMS, Dayton, Ohio, August 1998.Melodic Based Music Retrieval:
15.   A. Ghias, J. Logan, and D. Chamberlin. B. C. Smith, "Query By Humming." In Proc. ACM Multimedia 95, pp. 231-236, 1995.