Guojing, Sari and Hendry. Music Retrieval-By-Similarity Based Digital Music Management. International Journal of Business Society. Special 2021, 61-64

Journal Homepage: www.ijo-bs.com

International Journal of Business Society

Contents lists available at: https://www.ijo-bs.com/issue.html

MUSIC RETRIEVAL-BY-SIMILARITY BASED DIGITAL MUSIC MANAGEMENT

Ji Guojing[1]; Widya Sari[2*], Hendry[3]

[1]*Master student, Faculty of Management, Universitas Prima Indonesia*
[2]*Faculty of Management, Universitas Prima Indonesia, Email: (widyasari@unprimdn.ac.id)*
[3]*Faculty of Management, Universitas Prima Indonesia*
[*]*Corresponding Author*

Information of Article

Article history:
Received: 4 Nov 2021
Revised: 5 Nov 2021
Accepted: 30 Nov 2021
Available online: 1 Dec 2021

Keywords:
Digital music
Music management
Online music

ABSTRACT

The MIDI type of music has been extensively used in the research on music recovery to fine-tune acoustic difficulties based on the content. We are just interested in a conventional (MPEG layer three) online music archive's categorisation and recovery for this research. In this study, two quantitative tools are examined and rated. Gaussian Mixture Modelling works effectively in a music categorisation job with an accuracy of over ninety per cent. The vector quantisation approach that depends on the tree gets a somewhat insignificant impact, but the system is much quicker and more adaptable. Similarity-based music retrieving has also shown impressive outcomes, according to many researchers. Related coefficients are useful for describing the audio, but they take a long time to decompress. New parameterisations are consequently presented to permit the process of music at the interactive user rates based on the partial deconstruction of MPEG layer three audio. A typical computer music collection for management may benefit greatly from the strategies discussed here. Computer music collection for management may benefit greatly from the strategies discussed here.

1. Introduction

MPEG Layer 3 (MP3) and other unique codecs have made digital music libraries more ubiquitous in the past few years [1, 2]. Technologies for audio management will indeed be required to assist people in browsing and maintaining these collections as this tendency grows. Music categorisation and retrieving using techniques based on the content is the focus of this article. As a result, it adds to the scarce research on music retrieving based on the content.

Customers may utilise category information to narrow their search for music they like. It is fairly uncommon for the category comments in various compression formats to be erroneous or removed entirely. Text comments may be substituted or enhanced with user-defined subcategories that precisely represent the kind of song the listener is looking for. On the other hand, content-based music searching management may be of more use. The ability to identify new tracks that sound like a song or a brief excerpt in big (potentially web-wide) collections by retrieving sonically comparable music is an essential tool [3, 4]. Analysing purchase trends requires a lot of data and is only useful once the product is out [5]. The technology could be used in many different ways. It is possible to anticipate new music into one of two groups by keeping a customer's what to like and dislike [6].

As an example, digital radio broadcasts may be tailored to a listener's musical preferences by using this method [7]. In this work, two primary strategies for music processing are examined. The most often used method is Gaussian Mixture Modelling (GMM) when it comes to audio categorisation. The tree-Based Vector Quantisation (TreeQ) technique uses a racist and discriminatory approach and is explored [8]. The mathematical formulation of audio samples into extracted features is required for every strategy. It is common in the speech field to utilise Frequency Cepstral Coefficients. Another approach is shown that can bypass decompressing and parameterising the waveform of the compressed audio. This method, MP3/CEP, uses partly decoded MPEG layer three audio to quickly generate cepstral coefficients [9-12]. These approaches and parameters of the model are discussed in further depth in this work. Results of categorisation and retrieving management studies are then detailed and presented.

The paper is organised as follow: In Section 2, the applied methods and techniques. Section 3 discusses the results obtained from the experimental works. Section 4 draws the Conclusion for further future works.

https://dx.doi.org/10.30566/ijo-bs/2021.special.13

2. Applied Methods in this Research

A. Gaussian Mixtures Modeling (GMM)

Speaker identification is one of several audio text categorisation for which Gaussian Mixture Modeling has been effectively used [13]. As a probabilistic model of songs or music style, the audio and music management approach combines a combination of the probability of Gaussian element probability density.
This study may employ the mode since the texture features contain relatively uncorrelated components. The GMM parameters have been calculated such that the observations series of the training has the greatest likelihood of occurring. The probability may be regionally increased utilising the Expectation-Maximization approach, even though this problem cannot be solved in a boundary condition. Once the model parameters converge on the ultimate answer, this repeated pattern will repeat. A GMM can be used to estimate the likelihood that a testing feature map matches to the system. The full music may be composed by repeating all of the comments. Frames normalisation has been applied here instead of merely summarising this rating so that each frame has less impact.

B. Vector Quantization Using a Tree (TreeQ)

Second, a Vector Quantizer is trained instead of modelling the acoustics effectively. This approach has been used to identify speakers and retrieve brief sound and song samples. Extracted features are initially constructed from the parametric dataset. An artist or a category are connected with every training instance. A normalisation tree is built that dynamically divides the subspace into areas with the highest number of distinct classes. To create a histogram pattern for certain music or category, utilise the tree you just built. Every one of the tree's extracted features. A histogram pattern is created by comparing the number of examples at every cell to the total number of instances. Acoustics similarities may be estimated by comparing any two of these patterns, which reflect the music's acoustics in a compressed manner. This study uses the cosine distances metric to compare different matrices, despite other music management methodologies.

C. Schematics for Parameterisation

To begin with, converted to electronic music must be structured into feature vectors that may be used to process the audio. These vectors must maintain the most important information while removing extraneous acoustic information. To begin, MFCCs were used to evaluate the results [14, 15]. A thorough decompression of the sound is required. It is restored to a subband once the electronic waveform has been processed. Throughout parameterisation, although a distinct one, this partial duplication has been removed in the second approach, MP3/CEP, by employing relevant contents generated from the MP3 sub-band information.

3. Experimental Works

A. Classification of Musical Genres

Initially, the two approaches and parameters of the model have been evaluated to classify musical genres. These studies utilised music from a typical personal collection of MP3 files. Six musical subgenres with enough material for these attempts were identified for categorisation. Blues, relaxed listening, symphonic, opera, disco, and rocking are all examples of these categories. All the songs were divided equally into training and testing data. The testing set includes 150 songs from across all categories. Extra music has been utilised to build a general music genre or "trash model" to detect music that did not fit into any predetermined categories.

B. The GMM

The GMM technique for musical categorisation is evaluated by estimating a GMM from the training content for every genre. For each testing song, the finest GMM has been used to assign it to a certain genre category. A series of tests have been conducted to examine the performance of different MFCC input vectors configurations. An energy word is introduced in the second line after the fundamental cepstral elements are employed firstly. An energy component considerably increases the classification accuracy on the highly matching test. In addition, this could consistently improve accuracy. A decent performance may be attained with as few as four categorised elements, but the performance rapidly rises to a high of thirty-two elements. Ninety-five per cent of the 150 test songs are accurately identified using the most effective technique available today.

C. TreeQ Technique

There must be a single class for every soundtrack to apply tree-based vector quantisation. To build a quantification tree, the classes are used. Every category and testing file's leaf node elements are utilised to create a histogram pattern. The cosine distance measure has been used to classify every sample file and find the category with the nearest matched histogram pattern for every file. There are already comparisons between trees with 50, 200, and 400 leaves. An MFCC estimate can effectively depict temporal variability impacts by appending contextual factors into super-vectors since decision tree building is constant in the set of measurements. There have been four concatenation matrices that are used

https://dx.doi.org/10.30566/ijo-bs/2021.special.13

for categorisation. The findings show that a tree created utilising contextual and with two hundred leaves is the most successful classifier has an accuracy of the classification that equals to or greater than eighty-eight per cent.

D. Comparison of GMM and TreeQ Techniques

For the GMM and TreeQ techniques, it is summarised that the best findings from the results and outcomes is the top TreeQ score at eighty-three percentage points is somewhat less than the greatest GMM score at ninety per cent. Nevertheless, the computing expense of predicting and evaluating GMMs is a drawback. Despite efforts to alleviate this issue, such as accumulation, it continues to be a significant negative. Nevertheless, TreeQ's quantisation utilising the tree is very quick. It is also possible to use very effective searching patterns like M-trees [1] to discover related patterns. When velocity and flexibility are important, this technique is the best option.

E. Evaluation of Parameterizations

Tests employing MP3CEP selected features instead of MFCC can be observed. Similarly, GMM seems to be slightly surpassing TreeQ. For the finest GMM32 method, the classification accuracy with MP3/CEP settings is above ninety per cent, which is only one per cent lower than the leadership's MFCC value. This is a surprising finding, given that nearly twenty-two per cent fewer selected features were employed in this experiment. Performance drops from MFCC to MP3/CEP by 5 per cent for the TreeQ system. However, that might be a minor amount to be paid for the ability to analyse songs interactively. For one thing, MP3/CEP can only playback MP3 files. A mixed format file is not a good idea because of the lack of interoperability between different sound compression systems, e.g., MPEG2-AAC [16].

F. Music retrieval Performance

In a second trial, the two approaches were tested on a task that included retrieving music based on similarities. The data for this job has always been derived from that very same resource as the data for the categorisation test, which is described above. In every "soundtrack", 10 tracks have been selected and divided into two archives: a randomised 5 have been placed in a testing set, while the other five were placed in a series of potentially recoverable tracks. Several additional tracks were added to the retrievable candidate set to obscure the crucial ones, making it harder to complete this exercise. Approximately 200 tracks make up the final exam collection.

The findings suggest that song restoration is a pretty effective technique, and actual testing confirms these findings. It is uncommon to encounter instances in which two tracks are considered similar even though they sound nothing alike. It is important to note that the averaged accuracy values presented are inaccurate and unreliable depending upon the nature of the relevant requirements. However, they are useful in highlighting the differences between the different systems' performance requirements. When it comes to the tighter similar musician relevancy situation, GMMs outperform the TrecQ system significantly. Averagely, with an accuracy number of less than one, the most often recovered track has been one of the five possibilities by a single artist seventy per cent of the time, the best possible result. In the case of category significance, there is no statistically noteworthy variance between the two methodologies. More importantly, for audio recovery, the accuracy rate decreased by an average of four per cent throughout whole methods when the parameterisation was being used. As previously stated, this is tolerable for businesses that need fast action response times.

4. Conclusion

Using the findings presented in this work, it is shown that methods process and managing music methods may be used to create effective tools for navigating and maintaining online music archives. The performance of song categorisation is excellent, with the highest accuracy equal to eighty-nine per cent when differentiating between six different musical tracks. An accuracy value of less than one was also attained for music recovery in general, improving previous results. The free template matching approach outperforms GMM slightly among the strategies studied, even though the latter is the last and scalable solution. When used to user-interactive approaches, the MFCC parameterisation normally performs well irrespective of the music reduction management scheme.
Nevertheless, it is a bit sluggish when applied to music reduction schemes. An approach for parameterisation that is much quicker. As a result, the MPRCEP protocol was already suggested, based on sub-band content from partly decoded MPEG Layer three songs. It is expected that the mentioned strategies would find additional use in any sector that gives learning about a person's music tastes would be beneficial.

https://dx.doi.org/10.30566/ijo-bs/2021.special.13

References

[1] G. O'Regan, "MP3 Player and Digital Music," in The Innovation in Computing Companion: Springer, 2018, pp. 197-199.

[2] S. SECTOR and O. ITU, "Information technology–Generic coding of moving pictures and associated audio information: Video," ISO/IEC 13818-2, 2014.

[3] D. Yan, R. Wang, J. Zhou, C. Jin, and Z. Wang, "Compression history detection for MP3 audio," KSII Transactions on Internet and Information Systems (TIIS), vol. 12, no. 2, pp. 662-675, 2018.

[4] A. Louthander, "Perceived Audio Quality in Popular Music Encoded to MP3 and Opus," ed, 2019.

[5] L. U. Hasanah, T. W. Purboyo, and R. E. Saputra, "A Review of MP3 Steganography Methods," International Journal of Applied Engineering Research, vol. 13, no. 2, pp. 1128-1133, 2018.

[6] W. Rodgers, F. Yeung, C. Odindo, and W. Y. Degbey, "Artificial intelligence-driven music biometrics influencing customers' retail buying behavior," Journal of Business Research, vol. 126, pp. 401-414, 2021.

[7] G. Balasubramanian, A. Kanagasabai, J. Mohan, and N. G. Seshadri, "Music induced emotion using wavelet packet decomposition—An EEG study," Biomedical Signal Processing and Control, vol. 42, pp. 115-128, 2018.

[8] P. Kanawade and S. Gundal, "Tree structured vector quantisation based technique for speech compression," in 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI), 2017: IEEE, pp. 274-279.

[9] F. Maurer, S. Battista, L. Ciccarelli, G. Meardi, and S. Ferrara, "Overview of MPEG-5 Part 2–Low Complexity Enhancement Video Coding (LCEVC)," ITU Journal: ICT Discoveries, vol. 3, no. 1, 2020.

[10] J. K. Bii, "MPEG-1 Layer III Standard: A Simplified Theoretical Review," 2019.

[11] A. A. Hafiez and A. Rashad, "Theoretical Studies for Arsenic Interferences by Inductively Coupled Plasma Mass Spectrometer," 2019.

[12] J. Zhang, X. Yi, X. Zhao, and Y. Cao, "Light Multiscale Conventional Neural Network for MP3 Steganalysis," in Digital Forensics and Watermarking: 18th International Workshop, IWDW 2019, Chengdu, China, November 2–4, 2019, Revised Selected Papers, 2020, vol. 12022: Springer Nature, p. 43.

[13] I. Channoufi, S. Bourouis, N. Bouguila, and K. Hamrouni, "Image and video denoising by combining unsupervised bounded generalised gaussian mixture modeling and spatial information," Multimedia Tools and Applications, vol. 77, no. 19, pp. 25591-25606, 2018.

[14] Y. Li, X. Li, Y. Zhang, W. Wang, M. Liu, and X. Feng, "Acoustic scene classification using deep audio feature and BLSTM network," in 2018 International Conference on Audio, Language and Image Processing (ICALIP), 2018: IEEE, pp. 371-374.

[15] P. Doungpaisan and A. Mingkhwan, "Query by Example of Speaker Audio Signals using Power Spectrum and MFCCs," International Journal of Electrical and Computer Engineering, vol. 7, no. 6, p. 3369, 2017.

[16] M. Park and Y. H. Kim, "A Development of MPEG-2 TS-to-MMTP Stream Converter," Journal of Broadcast Engineering, vol. 25, no. 2, pp. 252-264, 2020.

https://dx.doi.org/10.30566/ijo-bs/2021.special.13