Audio Classification and Retrieval by Using Vector Quantization

Shruti Vaidya, Dr. Kamal Shah

Abstract—In today's world, we can say that information and its processing has become the critical aspect for functioning of everything. In the early days, information was generally obtained and processed in the form of text. Today information is available in all forms namely, text, music, graphics, etc. which are a easily understandable and accurately represent information. Information is first captured then the captured information is retrieved and analyzed for further requirements. In this paper, the information that we take into consideration is in audio form. We have studied the feature vector extraction methods, similarity measurement techniques, and have also measured the performance parameters. It has been observed that the use of multiple feature vectors provides better and more accurate classification and retrieval of audios from large database.

Index Terms— Audio, Audio Retrieval, Audio Vector Quantization, Data Compression, k-Nearest Neighbor, Precision Recall, Vector Quantization

1 Introduction

Vector Quantization (VQ) is an efficient and simple approach for data compression. Since it is simple and easy to implement, it is widely used in different applications, such as pattern recognition, face detection, image segmentation, speech data compression, Content Based Image Retrieval, tumor detection etc. Vector quantization is a lossy compression technique. There are three major procedures in vector quantization, namely codebook generation, encoding procedure and decoding procedure. In the codebook generation process, audio is divided into several k-dimensional training vectors. The representative codebook is generated from these training vectors by the clustering techniques. In the encoding procedure, the original audio is divided into numerous k-dimensional vectors and the encoding of each vector is done by indexing of codeword by a look up table methodology. The encoded results are called an index table.

In the decoding procedure, the same codebook is used by the receiver to translate the index back again into its appropriate codeword for rebuilding of the audio. One of the key points of Vector Quantization is to generate a good codebook such that the distortion between the original and the reconstructed audio should be minimum. In order to find the best-matched codeword in the encoder, various codebook full search algorithm can be used [1].

2 Overview of System

Research till today in audio classification tends to focus on matching test sounds into a limited number of predefined categories such as music, applause, speech etc., but this approach would describe each sound on the feature vectors.

Furthermore, the proposed system allows intelligent interpretation of unseen examples, e.g. describe a door closing based on the similarity to previously seen events. The new signal can be easily classified and other related sounds can also be retrieved in relation to the other sounds as shown in the Fig.1.

For instance, consider a system where given an input sound of a door closing, would return the label "background sound", and will retrieve from a database samples most similar to it.

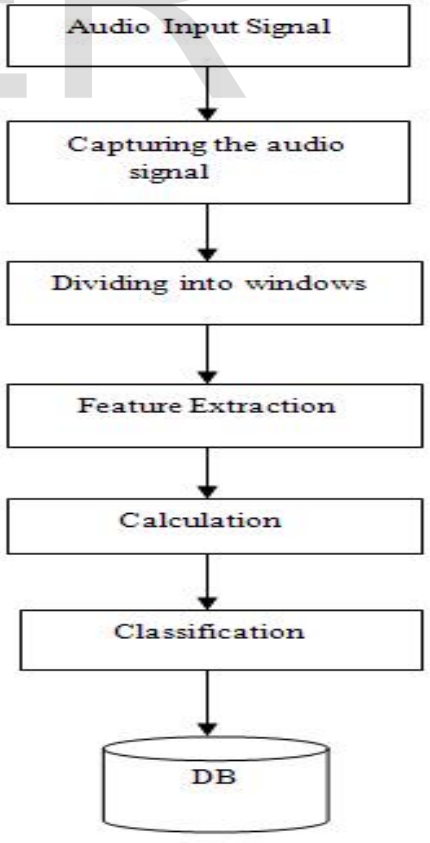

Fig.1: System Overview

- *Shruti Vaidya is currently pursuing masters of engineering degree program in information technology, TCET, Mumbai University,India, E-mail: shrutiv01@gmail.com*
- *Dr. Kamal Shah is currently a professor in masters of engineering information technology department,TCET, Mumbai University,India, E-mail: kamal.shah@thakureducation.org*

3 AUDIO RETRIEVAL

Music retrieval for our system undergoes two steps: It undergoes feature extraction process after which appropriate processing is done.

3.1 Feature Vector Extraction

Feature Extraction is the process of converting an audio signal into a sequence of feature vectors which carries characteristic information about the signal. These vectors are used as the basis of various types of algorithms used for audio analysis is based on the computation of features on the basis of windows. These window based features can be considered as short time description of the signal for that particular moment in time. The performance of a set of features depends on the application. A wide range of audio features exist the purpose of classification. These features belong to two categories: time domain and frequency domain features [2].

We consider the following multiple features [3],[4]:-

- Spectral Flux

The spectrums average variation value, in one second window, between two uninterrupted frames is recognized as Spectrum Flux. Spectrum Flux provides good mark of demarcation between speech, sound and environment.

- Low Short Time Energy Ratio

Variation of short time energy is measured by low short time energy ratio. The relative amount of the number of frames, where the short term energy is not more than 0.5 times of the average short time energy is referred as Low short Time Energy Ratio.

- Zero Crossing Rate Ratio

Zero-crossing rate (ZCR) is proved to be useful in classifying different audio signals. It has been more widely used in speech/music classification algorithms. According to observed results, it is seen that the variation of zero crossing is more distinguishing than its exact value. Hence, high zero-crossing rate is used. The ratio of the numerous frames where the zero crossing is above the 1.5 fold average zero crossing, is recognized as High Zero Crossing Rate Ratio.

- Spectral Centroid

Spectral Centroid belongs to frequency domain feature vector. The centre of gravity of the magnitude spectrum of the STFT is defined as the spectral centroid.

- Spectral Roll Off

Spectral Roll Off is also a frequency domain feature vector. The spectral roll off is defined as the frequency Rt, where 85% of the magnitude distribution is concentrated under.

3.2 Method Used

There are many techniques available that can be used. We will be considering the following algorithms:

1. Fast Fourier Transform

Fast Fourier Transform (FFT) is an algorithm used for computation of Fourier transform as well as its inverse. It converts from time domain to frequency domain. For later processing, after the conversion, feature vectors from the frequency domain can be applied.

2. k-Nearest Neighbor

This method consists of assigning to the unlabelled feature vector or the label of the training vector that is nearest to it in the feature space. In KNN a training set T is used to determine the class of a previously unseen sample X. A suitable distance is measured in the feature space between the unseen sample and all the samples of the training data. This distance is used to determine k element in T closest to X. and if most of these k nearest neighbors contain similar values, then X gets classified accordingly. These classification schemes clearly define non-linear decision boundaries and thus improve the performance [5], [6].

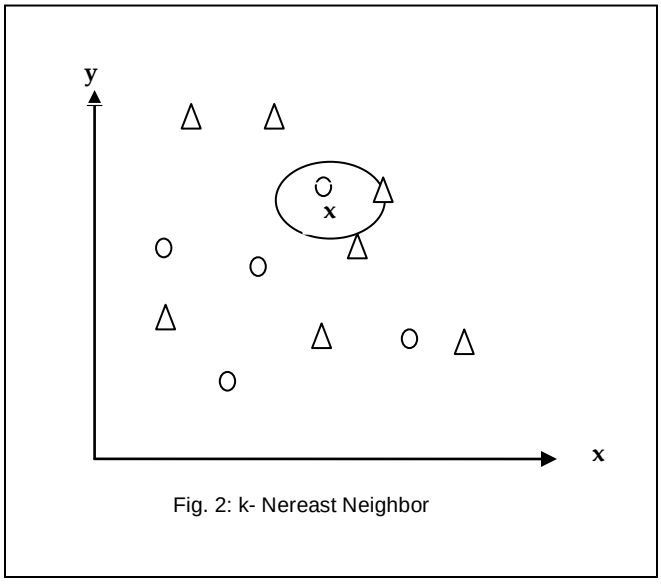

Fig. 2: k- Nereast Neighbor

4 Distance Measure

Once the feature vectors are applied and the domain of the unknown signal is determined, retrieval of similar signals needs to be done. Distance calculation provides support. We consider the Euclidean Distance and Manhattan Distance for the purpose of retrieving the similar signals [7].

- Manhattan Distance

The Manhattan Distance obtains the distance when moved from one place to another when a grid path is followed. It is the sum of the differences of their corresponding elements.

- Euclidean Distance

The Euclidean Distance can be obtained when moving at an angle of 45 degree. It is the square root of the differences between corresponding elements.

5 Performance Parameters

We present our on different audio input signals for different categories. The Precision provides the accuracy. Recall provides us with accurateness [8].

$$\text{Precision} = \frac{\text{Number of relevant audios retrieved}}{\text{Total number of audios retrieved}} \quad (1)$$

$$\text{Recall} = \frac{\text{Number of relevant audios retrieved}}{\text{Total number of relevant audios in database}} \quad (2)$$

The above equations provide us with the values of Precision and Recall. These are then plotted against the x-y axis. The point where they meet is called as the crossover point. This crossover point gives us the performance measure. The higher the value the more efficient.
The Precision Recall Graph for some categories are as follows:-

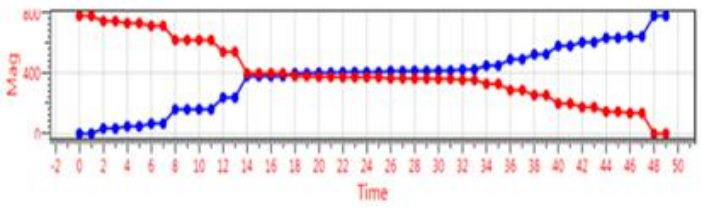

Fig. 3: Precision Recall Graph for Speech [9]

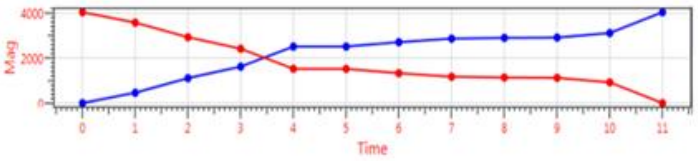

Fig. 4: Precision Recall Graph for Background [9]

6 Conclusion

We have presented our study on unknown audio signals taken in wave formats. The unknown was effectively classified into its appropriate category by means of the various feature vectors taken into account. We also used Fourier Transform for conversion from time domain to frequency domain, as also the k-nearest neighbor classifier. With the unknown signal being identified into its appropriate category, the system also retrieved the other signals belonging or similar to the new signal arrived. Performance parameters provided efficient and satisfactory results.

7 References

[1] H. B. Kekre, Tanuja K. Sarode And Jagruti K. Save, "Error Vector Rotation Using Kekre Transform For Efficient Clustering In Vector Quantization" International Journal Of Advances In Engineering & Technology, July 2012.©Ijaet Issn: 2231-1963

[2] Vaishali Nandedkar, Computer Dept.,JSPM,Pune, " Audio Retrieval Using Multiple Feature Vectors", International Journal of Electrical and Electronics Engineering (IJEEE), Volume-1, Issue-1, 2011

[3] Lie Lu, Hong-Jiang Zhang, Senior Member, IEEE, and? Hao Jiang, "Content Analysis for Audio Classification and Segmentation" , IEEE Transactions on Speech And Audio Processing, Vol. 10, No.7, October 2002.

[4] Cheng-ya Sha, "Time Frequency Analysis for Acoustics" ,ntu.edu.tw.

[5] Hariharan Subramanian, Prof. Preeti Rao, Dr. Sumantra. D. Roy "Auido Signal Classification" M.Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay, Submitted November 2004.

[6] Finnish Meteorological Institute, www.geo.fmi.fi/~syrjasuo/Analysis/node6.html, October 2004.

[7] T. Soni Madhulatha, Associate Professor, Alluri Institute of Management Sciences, Warangal, " An Overview On Clustering Methods" , IOSR Journal of Engineering Apr. 2012, Vol. 2(4) pp: 719-72

[8] H.B.Kekre, Sundeep D. Thepade, Tanuja K. Sarode, Shrikant P. Sanas, " Image Retrieval Using Texture Features Extracted Using LBG, KPE, KFCG, KMCG, KEVR with Assorted Color Spaces" , International Journal of Advances in Engineering & Technology, Jan. 2012. ISSN: 2231-1963

[9] Shruti Vaidya, Dr. Kamal Shah, "Application of Vector Quantization for Audio Retrieval", International Journal of Computer Application (0975-8887) Volume 88- No.17,February 2014, pp.23-27