iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** VIII **Month of publication:** August 2015

DOI:

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Classification of Audio Signal Using Local Discriminant Bases

S. V. Dhabarde[1], Prof. P. S. Deshpande[2]
[1]*Department of EnTC, University of Pune, India*

Abstract— In the age of digital information, audio data has become an important part in many modern computer applications. Audio classification has been becoming a focus in the research of audio processing and pattern recognition. Automatic audio classification is very useful to audio indexing, content-based audio retrieval and on-line audio distribution, but it is a challenge to extract the most common and salient themes from unstructured raw audio data. To generate the highest quality classification of audio signal. To obtain high accuracy for this classification into natural and artificial classes in this Local Discriminant Bases (LDB) technique the main aim here is to analyze and characterize audio content and the key feature is to identify discriminatory subspaces using LDB technique. To use simple features for the classification and reduce the complexity of the system.
Keywords— Feature Extaction; database; LDB technique; tree classification.

I. INTRODUCTION

Audio feature extraction has a very large importance in our day to day life. Consider an example: people with hearing disability depend on assistive devices such as hearing aids to listen to the sounds around them. It is very important for these assistive devices to determine the environment using the auditory clues in order to build better instruments with automatic switching features. This would improve the quality of life of people with disability. To consider another field of relevance considers audio environment detection would be in wildlife conservatories. Animal, bird, and insect sounds can be used to automatically keep track of the wildlife population and their movement patterns. By using audio-triggered gadgets, it would be possible to understand and create unique audio patterns representing the life style of various species. In order to perform applications such as the one mentioned above, efficient audio feature extraction and classification schemes covering wider categories of audio signals are required.

The main objectives of this project have been mentioned below,

To generate the highest quality classification of audio signal.

To obtain high accuracy for this classification into natural and artificial classes

In this Local Discriminant Bases (LDB) technique the main aim here is to analyse and characterize audio content and the key feature is to identify discriminatory subspaces using LDB technique.

To use simple features for the classification and reduce the complexity of the system.

This chapter introduces the main concept of what cervical cancer is and how it is dangerous if not cured in time. It even explains us with advanced automation; data is evaluated in real time to extract the information, which is further used for medical application.

II. METHODOLOGY

The proposed methodology is an attempt to develop local discriminant bases (LDBs)-based automated multigroup audio classification system. The proposed methodology addresses only the sound source (type) identification part of an auditory scene analysis and not the complete auditory analysis by itself. Audio signals are highly no stationary in nature, and the best way to analyze them is to use a joint time-frequency (TF) approach. The previous works [3], [4] of the authors have demonstrated the success of the TF approach in music signal classification. In order to perform efficient TF analysis on the signals for feature extraction and classification purposes, it is essential to locate the subspaces on the TF plane that demonstrate high discrimination between different classes of the signals. Once the target subspaces are identified, it is easier to extract relevant features for classifications. In the proposed work, we used LDB [3] with wavelet packet bases to identify these target subspaces in the TF plane to classify the audio signals. The optimal choice of LDBs depends on the nature of the dataset and the dissimilarity measures used to distinguish between and among classes. A combination of multiple dissimilarity measures can be used to achieve high classification

Volume 3 Issue VIII, August 2015
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

accuracies. The proposed work follows the nonstationary signal analysis approach similar to our previous work [3]; however, compared to our previous work, the proposed work focuses on developing a generic methodology for automatic identification of discriminatory subspaces and uses simple features to classify a wider range of audio signals. The proposed work will also be computationally less expensive as the orthonormal wavelet packet basis sets are much smaller than the overcomplete and redundant TF dictionaries used in our previous study.

A block diagram of the developed system is shown in Fig. 3.2. Here input is taken in the form of audio database .Therefore select a random audio signal as input consisting of both, natural (human, birds, etc) as well as artificial (music, cars, machines,etc).Training signal pair wise comparison should be done. For that we have to divide Audio signal into classes.

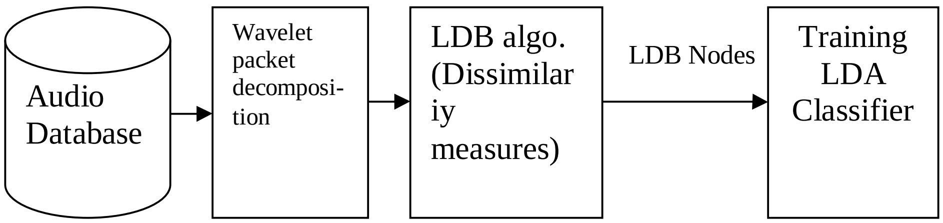


Fig 1. Block diagram of Training phase

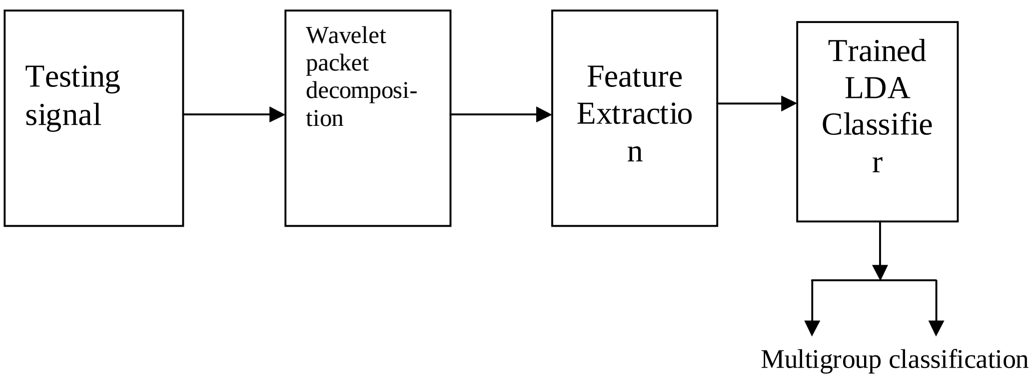


Fig 2. Block diagram of testing phase

Algorithm

Step 1. Input is taken in the form of audio signal for training of the system.

Step 2. This training signal is decomposed using wavelet packet decomposition.

Step 3. Then Local discriminant bases(LDB) is applied to decomposed packets till the system is completely trained.

Step 4. LDB nodes are created inorder to test the testing inputs.

Step 5. The feature extaction and Linear Discriminant Analysis(lda) classification of the test signal is done.

Step 6. Output in the form of multigroup classification is obtained.

III.IMPLEMENTATION

In LDB algorithm, an audio signal *xi* is taken as an input and decomposed into a binary wavelet packets tree *{Ωj;k}*, where *j* indicates the level of the tree, *k* represents the node index in level *j*, *Ωj;k* denotes the subspace spanned by a set of wavelet packet basis vectors $\{W_{j,k,l}\}_{l=0 \text{ to } 2^{(u-j)}-1}$ and 2^*u* corresponds to the length of the signal *xi*. So the signal *xi* can then be given as

$$x_i = \sum_{j,k,l} [\alpha_{j,k,l}]_i . w_{j,k,l} \qquad (1)$$

Where, $a_{j;k;l}$ are the basis vector coefficients at node (*j; k*).

The main aim of LDB is to identify the set of subspaces (nodes), which provides maximum dissimilarity information between the

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

different classes of the signal, from all 2*J* mutually orthogonal subspaces (*J* is the number of decomposition levels). This is a pruning process of the wavelet packet tree based on the discriminative capability of a wavelet packet node (subspace) (*j; k*) on *P* signal classes, for which following steps are adopted:

Step 1: Calculate the dissimilarity values $M_\beta(j;k)$ of node (*j; k*) on the $\beta^{th} \in \binom{p}{2}$ combination of two signal classes, which is given as the mean of the dissimilarity values $\{D^{(j;k)}\}_{t \in [1..T]}$ between a set of *T* sample signal pairs taken randomly from the training sets of two classes. As in [6], $D^{(j;k)}$ is computed as the difference of the normalized energy, $E_1^{(j;k)}$-$E_2^{(j;k)}$, or the difference of temporal variance of local piecewise variances of wavelet packet basis coefficients, $\frac{\mathrm{var}[v^1_{(j,k)}]}{\mathrm{var}[v^2_{(j,k)}]}$, on node (*j; k*) of two classes samples being Compared, which reveal the variance of energy concentration locations on the TF plane or the non-stationary for these two audio classes.

Step 2: Choose *Q* LDB nodes with consistently high values of $M_\beta^{(j;k)}$over the random trials of training samples.

Step 3: From the group of $Q * \binom{p}{2}$ LDB nodes collected from all class combinations, choose the *H* most frequently occurring LDB nodes, for which we can extract the wavelet packet basis coefficients or the energy or variance value to form the final feature vectors.

A. LDA Approach in 5 Steps

Listed below are the 5 general steps for performing a linear discriminant analysis,

Compute the *d*-dimensional mean vectors for the different classes from the dataset.

Compute the scatter matrices (between-class and within-class scatter matrix).

Compute the eigenvectors ($\mathbf{e_1}, \mathbf{e_2}, ..., \mathbf{e_d}$) and corresponding eigenvalues ($\lambda_1, \lambda_2, ..., \lambda_d$) for the scatter matrices.

Sort the eigenvectors by decreasing eigenvalues and choose **k** eigenvectors with the largest eigenvalues to form a $d \times k$-dimensional matrix **W** (where every column represents an eigenvector).

Use this $d \times k$ eigenvector matrix to transform the samples onto the new subspace. This can be summarized by the equation $\boldsymbol{Y} = \boldsymbol{X} \times \boldsymbol{W}$ (where $\boldsymbol{X}$ is an $n \times d$-dimensional matrix; the *i*th row represents the *i*th sample, and $\boldsymbol{Y}$ is the transformed $n \times k$-dimensional matrix with the *n* samples projected into the new subspace) Audio feature extraction has a very large importance in our day to day life. Consider people with hearing disability depend on assistive devices such as hearing aids to listen to the sounds around them. It is very important for these assistive devices to determine the environment using the auditory clues in order to build better instruments with automatic switching features . This would improve the quality of life of people with disability. To consider another field of relevance consider audio environment detection would be in wildlife conservatories.

IV. RESULTS AND DISCUSSION

The correct classification accuracy in percentage for each of the classes is presented using following methods, which is regular LDA (resubstitution method). In the regular LDA method, the LDA-based classifier was trained with all available data and tested with all available data. This is called the resubstitution method of estimating classification accuracy . Since all the data is used for training and testing, this method is optimistically biased and, hence, the results can be considered as the upper bound of the achievable accuracy. The results are obtained using software MATLAB.

Volume 3 Issue VIII, August 2015
ISSN: 2321-9653

INPUT FOR NATURAL SOUND OF A PEACOCK

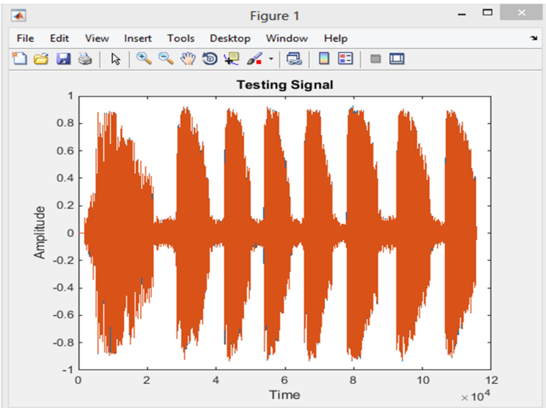


Fig 6.1: Input for natural sound of bird.

OUTPUT

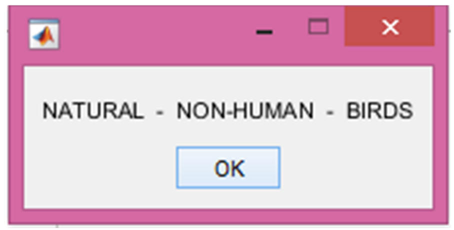


Fig 6.4: Massage box for output.

INPUT FOR ARTIFICIAL SOUND OF A TRAIN

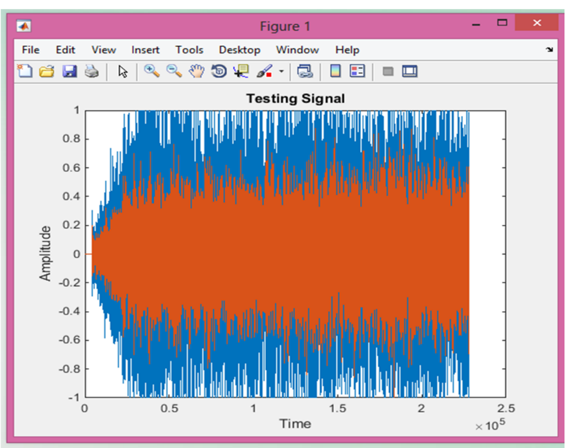


Fig 6.1: Input for Artificial sound of train.

OUTPUT

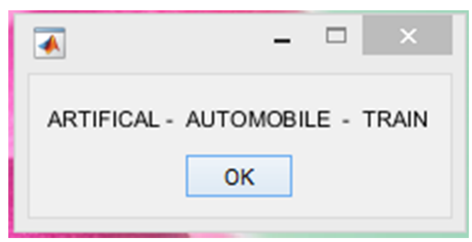


Fig 6.4: Massage box for output.

Table 6.1: Result table using LDB algorithm.

First level	First level
Natural 93%	Artificial 95%
Second level	Second level

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Human 95%		Non-human 90%		Instruments 97%		Automobiles 95%	
Third level				Third level			
Male 95%	Female 95%	Animals 85%	Birds 75%	Drums 97%	Piano 97%	Aircrafts 96%	Helicopters 95%

V. CONCLUSION

LDB algorithm helps in achieving high classification accuracy. LDB also uses simple dissimilarity measures like node energy and non-stationary index which perform well in identifying the discriminatory nodes between audio classes. Also LDB based audio classification has a significant potential in auditory scene analysis or environment detection. LDB performs good for all signals while MFCC works well for music. Combination of MFCC and LDB also gives promising results

VI. ACKNOWLEDGMENT

I would like to thank my beloved and respected guide Prof. Pallavi. S. Deshpande for her encouragement and support. I gratefully acknowledge her for imparting me valuable basic knowledge of Speech Processing. We are also thankful to the department of Electronics and Communication, Smt.Kashibai Navale College of Engineering, Pune for providing us infrastructure facilities and moral support.

REFERENCES

[1] K. Umapathy, S. Krishnan and R. K. Rao, "Audio signal feature extraction and classification using Local discriminant bases ," IEEE Trans. On Audio,Speech and Language Processing ,vol. 15, no. 4, pp.1236-1246, Mar. 2007

[2] R. Cai,L. Lu, A. Hanjalic, H.-J. Zhang and L.-H. Cai, "A Flexible framework for key audio effects detection and auditory context inference," IEEE Trans. On Audio, Speech, and Language Processing, vol. 14, no. 3, pp. 1026-1039, May. 2006.

[3] N. Saito and R. R. Coifman , "Local discriminant bases and their applications," Journal of Mathematical Imaging and Vision, vol. 5, no. 4, pp. 337-358,1995.

[4] K. Umapathy, S. Krishnan, and S. Jimaa, "Multigroup classification of audio signal using time-frequency parameters" IEEE Trans. On Multimedia , vol. 7, no. 2, pp. 308-315, Apr. 2005.

[5] Olivier Lartillot, Petri Toiviainen, "A MATLAB toolbox for musical feature extraction from audio,"Proc. Of 10th Internationl Conference on Digital Audio Effects(DAFx-07) , sept. 2007.

[6] G. Tzanetakis and P. Cook, "Marsyas: A framework for audio analysis," Organized sound , vol. 4, no. 3, 2000.

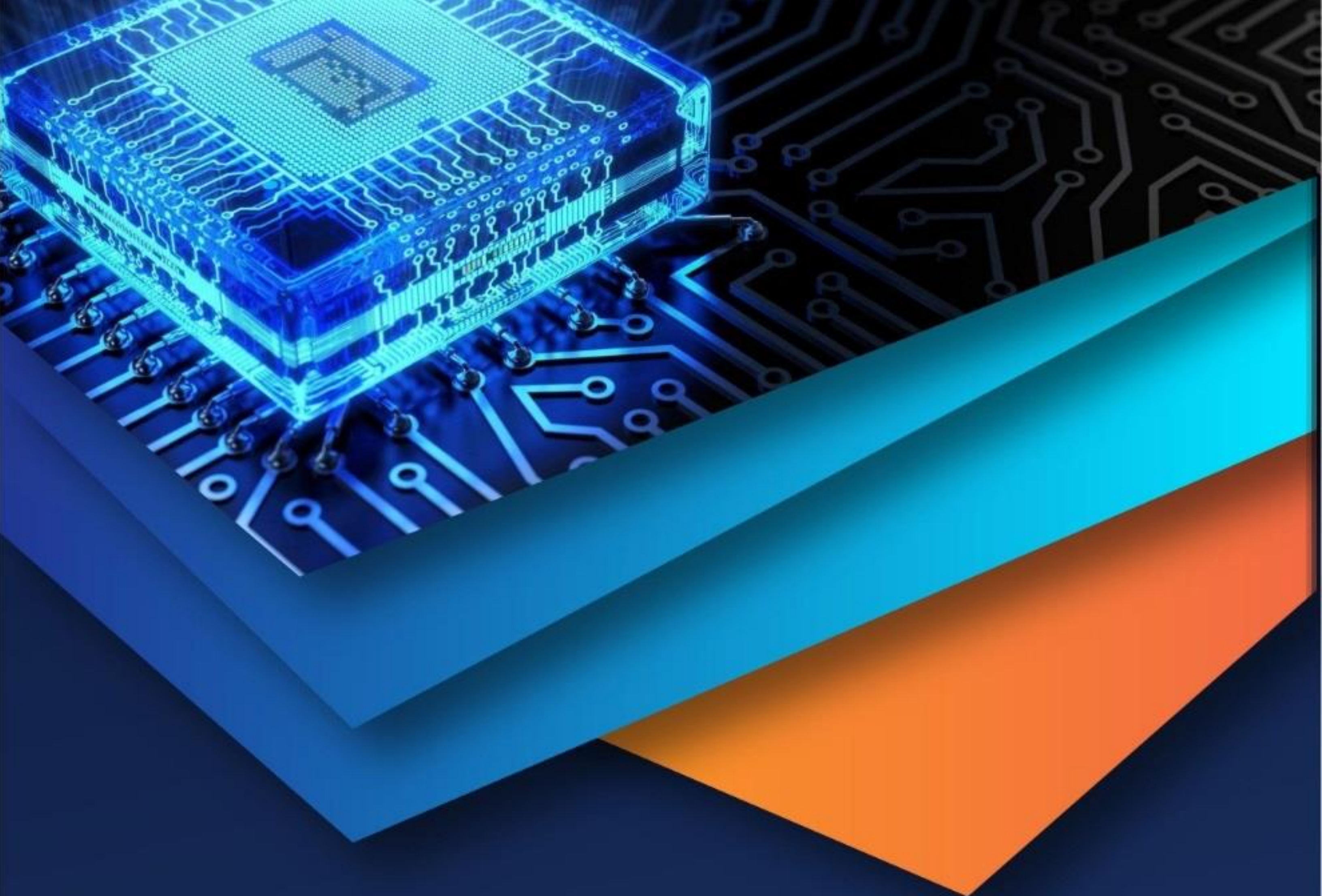
[7] M. A. Casey, R. Veltkamp, M. Goto,M. Leman,c. Rhodes, and M. Slaney, "Content based music information retrieval: current directions and further challenges,"Proceeding of the IEEE, vol. 96,no. 4, pp. 668-696, Jan. 2008.

[8] G. Wichern,J. xue,H. Thornburg,B. Mechtley, and A. S. Spanias,, "Segmentation,Indexing,and retrieval of environmental and natural sounds"IEEE trans. On Audio, Speech and Language Processing , vol. 18, no. 3,pp. 688-707,2010.

[9] C. Xu, N. C. Maddage, and X. Shao"Automatic music classification and summerization,"IEEE transaction on speech audio processing,"vol. 13, no. 3, pp. 441-450, May 2005

[10] Selina Chu, S. Narayanan, and Jay Kuo, "Environmental sounds recognition with time-frequency audio features," IEEE transaction on Audio, Speech, and Language Processing , vol. 17, no.9 , Aug. 2009

[11] K. M. Indrebo, R. J. Povinelli, M. T. Johnson, " Minimum mean-squared error estimation of Mel-Frequency Cepstral Coefficients using a novel distortion model," IEEE transaction on Audio, Speech, and Language Processing , vol. 16, no.8 , Nov. 2008



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY