International Journal of Soft Computing and Engineering (IJSCE)
ISSN: 2231-2307, Volume-2 Issue-4, September 2012

Movie Piracy Detection Based on Audio Features Using Mel-Frequency Cepstral Coefficients and Vector Quantization

B. Srinivas, K.Venkata Rao, P. Suresh Varma

Abstract— ***Along with the increase in the advancement of technology in movie industry over internet, there is also an increase in the movie piracy via internet which affects factors like economy and repudiation of movie industry. Internet movie piracy is the most common means for pirates as well as downloader's to break copyright laws by anonymous illegal uploads/downloads. In this paper we proposed an automated internet movie piracy detection mechanism based on audio fingerprint, which implements two famous algorithms, one is Mel-Frequency Cepstral Coefficients (MFCC) for feature extraction and the other is Vector Quantization (VQ) for classification. Our trained system initially looks for the sites which offer illegitimate copies of movies and if there is any suspicion based on a particular movie which is similar to the database of copyrighted movies that are registered with our trained system, it simply compares the fingerprints that are generated by implementing the above specified algorithms for both the trained and suspected movies. We collected various audio samples of different movies and we also extracted audio samples of pirated movies via internet with and without noises and trained and tested with our system. Finally, our system rendered efficient results with few error rates. We collected 52 audio samples without noise and 48 samples with noise and the resulted success classification is 96% and 92% respectively.***

Index Terms— ***Classification, Code Book, Movie Piracy, MFCC, VQ.***

I. INTRODUCTION

Movie piracy is committed in many ways, including via the internet by downloading and swapping movies, and on the streets, where illegally duplicated VCDs and DVDs are sold by shop owners and street vendors. Any person who sells, acquires, copies or distributes copyrighted materials illegally is called as a pirate. These pirates form groups which are generally known as scene groups or warez groups to release movies in several formats and in different versions via internet.

The digital age has ushered in a plethora of ways to steal movies via internet. Downloading a movie without paying for it is morally and ethically indifferent from walking into a store and stealing a DVD off the shelf. This hurts a country's economy, consumers and movie industry. The Motion Picture Association of America (MPAA) estimates that "a major movie on location contributes around $200,000 per day to local economies" due to studio employees using local businesses. Piracy detracts from these economic relationships. Distributors, theaters, video rental companies and pay-per-view providers all lose money.

It is the responsibility of the marketing department or an individual producing department to take action against piracy for their movie. However, catching these bootlegged and counterfeit movies via internet takes thousands of man-hours to watch the websites which offers illegal movies. To overcome these problems automated systems or servers were introduced, which works continuously to detect pirated movies through internet all over the world. There are many automated techniques which are introduced to detect these pirated movies according to their requirements. But some of these automated processes faced problems when the video has been altered or when a noise has been added.

In this paper we proposed a novel approach to stop pirates, which uses an audio recognition technique to detect pirated movies. The main idea is to develop an audio similarity system to retrieve and classify audio files based on their acoustical content such as loudness, pitch, bandwidth and harmonistic properties. Initially any person who wants to protect their movie against piracy has to register with our system and should produce appropriate audio samples or the audio of entire movie. Our system will then parameterize each audio into their features, which are used as signatures to uniquely identify each file. Our system then scans the internet to detect websites which offers pirated movies, when there is a suspicion regarding any content our system simply extracts the audio and computes its fingerprint and compare this to a database of fingerprints. Comparing these fingerprints rather than audio is more efficient and effective because irrelevant data have been removed in the process of computing the fingerprint, resulting in improved data search rate. If these finger prints are similar then the suspected movie is considered as a pirated one. Our system is efficient for detection even if there are disturbances or noises or even the other transformations are added.

Our system implements two main methods, one is "feature extraction" and the other is "feature matching". We used MFCC algorithm to simulate feature extraction module, using this algorithm the cepstral coefficients are calculated on mel frequency scale. VQ (Vector Quantization) method is used to reduce the amount of data in the form of codebooks in order to decrease computation time. These data are saved as acoustic vectors in the database. In the feature matching stage the features of suspected movie is compared against the database of signatures.

Manuscript received September 02, 2012.

B.Srinivas, Computer Science and Engineering Department, M.V.G.R College of Engineering, Vizianagaram, India.

Prof. K.Venkata Rao, Head of The Department, Computer Science and Engineering,Vignan IIT, Vishakapatnam, India.

Prof. P.Suresh Varma, Principal, Professor, Computer Science and Engineering, University College Adikavi Nannaya University, Rajahmundry, Andhra Pradesh, India.

II. RELATED WORK

Our solution to detect movie piracy mainly focus on audio recognition process using famous MFCC and VQ algorithms. So, our study mainly focused on the functionality and feasibility factors of MFCC and VQ algorithms, and then we studied different movie piracy detection techniques.

E.G.Rajan et al [1], proposed a system for speaker recognition using MFCC and VQ, they considered different factors that are affecting the performance of speaker recognition like noise, text dependencies, quality of voice samples, length of each sample etc., and their detection criteria is purely based on the error rate of these specified factors. If the error rate is below 10% between the training and testing vectors then their system recognizes that the training and testing samples belongs to the same speaker. Lindasalwa et al [2], discussed MFCC and DTW algorithms which are important to improve the performance of voice recognition process, their technique will authenticate particular speaker based on individual information that was included in voice signal, and tested these signals of particular speaker with other signals by adding different factors. And their results proved that their proposed technique is very efficient for recognition. Satyanand singh et al [3], proposed and evaluated a vector quantization approach for speaker recognition using MFCC and inverted MFCC for text independent speaker recognition. Their idea for speaker recognition mainly focused on inverted MFCC, some features of speaker which were ignored by MFCC will be captured by inverted MFCC and the output resulted will be modeled by VQ. They stated that their technique is very effective in real world implementation for speaker recognition.

Michael A. Lewis et al [4], they initially stated a problem, that is, when two equal bandwidth signals are added by one or two speakers, then the separation of individual speakers co-channel signal is not possible. Later they came up with a solution by designing a new scalar feature called Pitch Prediction Feature (PPF), this outperforms the linear predictive cepstrum (LPCC) and mel-wrapped cepstrum (MFCC) and then classified these features by VQ and neural tree network (NTN) individually. Their comparison for classification stated that VQ is more consistent than NTN by considering the lowest code book size using PPF i.e.1. Shi-Huang Chen et al [5], used Mel-Frequency cepstral coefficients (MFCC) and Support Vector Machine (SVM) to the task of text-dependent speaker verification system [5, 14]. Their technique initially applies feature extraction using MFCC from the voiced password provided by the user. Then they implemented SVM to train the speaker characteristics model from these MFCCs. Laura E. Boucheron et al [6], proposed a method for low bit-rate coding of speech through VQ of mel-frequency cepstral coefficients. Their proposed MFCC based codec has mean opinion score along with this their proposed codec can better preserve the fidelity of MFCCs features which can be further used as features of DSR for better word accuracy rate.

Our proposed mechanism mainly focuses on detecting pirated movies via Internet. These pirated movies can be offered in the internet using three different methods i.e. by IRC (internet Relay Chat), by P2P (peer to peer) networks like bit torrents and finally by FTP. Due to the immense increase in the growth of internet and easy navigation, people often download free illegal copies of movies from sharing networks, pirate websites, warez group sites, or from servers offering to download buccaneered copies of legitimate movies. Detecting movie piracy over internet is very difficult task because it involves thousands of man hours for detection. To overcome this problem automated systems were designed to scan the entire internet in order to look for the web sites that are offering pirated movies. In this paper our proposed mechanism is also an automated technique to gather audio bits from the suspected movie from the web sites offering bootlegged movies and to compare these fingerprints against our trained system database.

Some of the previous study and automated detection techniques on movie piracy over internet are as follows: Kristleifur et al [7]. Proposed a mechanism using Videntifier, they considered two scenarios whereas in the first scenario, Videntifier detects pirated copies from internet even they are modified and then these videos are decomposed into frames using Eff^2 descriptors. In the second scenario they proposed video monitoring system in order to detect movie piracy over the internet. Marc Fetscherin [8] presented a model by considering various variables to understand the customers behavior, these variables includes economic (income), demographic (age, gender), risk (virus), cultural facorts(education) etc., they had also assessed the quantity and quality of movie copies on peer-to-peer networks by considering most known file sharing system KaZAa [8]. Finally their survey resulted that many movie copies in peer to peer networks are not copyrighted and the probability of high quality movie copies are very less. Their assessment on the behavior of various customers also proved that many people are often downloading pirated copies.

III. METHODOLOGY

In this paper we proposed a novel technique to detect internet movie piracy. For each and every movie the fingerprint of the audio content which includes dialogues, background music etc., are very unique. With this idea we proposed an efficient method to detect movie piracy using audio content similarity by extracting audio features and by comparing these features. If these features are similar then the suspected movie is considered as pirated movie.

It is the responsibility of an individual person who involves in making a film, to protect his film against piracy. In our system one need to register their movie in order to detect piracy. These registered movies audio samples or the entire audio content of the movie will be then extracted and trained with our system using two efficient algorithms called MFCC and VQ. And the signatures of these trained audio contents are saved in the database. Our server will then look over the internet to search for the sites which offers pirated movies against copyright act. If the system finds such kind of sites it simply downloads the suspected movie and implement the above specified algorithms to its extracted audio content in order to generate signatures. These signatures are then compared against the database and if our system finds any similarity then the suspected movie is considered as pirated one.

The most important steps of training and testing sessions of our system involve feature extraction and classification method.

The aim of feature extraction step is to strip unnecessary information from the entire data and to convert the audio properties which are important for pattern recognition to a simplified format for distinction of the classes. This process reduces the dimension of the data in order to avoid curse of dimensionality. The aim of classification is to estimate general extension of the classes within feature space from a training set.

3.1 Feature Extraction

The audio content in this step must be processed to extract important characteristics. We used Mel frequency cepstral coefficients (MFCC) for feature extraction. The pattern recognition process cuts the digitized signal i.e. the sequence of sample values into overlapping windows of equal length. These cut-out portions of the digital audio signal are called as frames. Sometimes, for audio recognition tasks longer frames are used in comparison to feature extraction method which will be used for recognition process in order to increase spectral resolution.

Each extracted frame is then transformed into a MFCC vector. Therefore, extracted original signal is converted into a sequence of feature vectors, each representing cepstral properties of the signal within corresponding window. Here "a" is between 0.9 and 1.

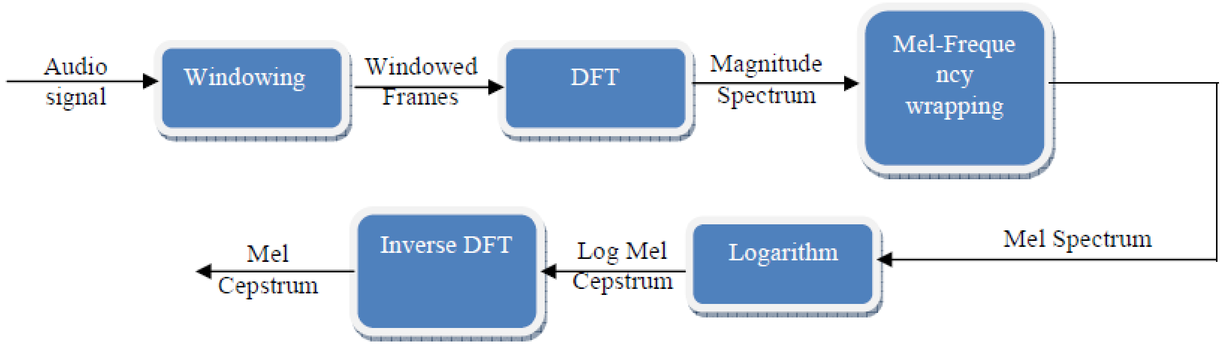


Fig 1. MFCC Computation Computation

3.1.1 MFCC Computation

The MFCC processor involves the following steps

Step-1: The continuous audio signal is blocked into frames of N samples, with each adjacent frames being separated by M (M<N).

Step-2: This step involves in widowing each individual frame so as to minimize the signal discontinuities at the starting and the end of each frame. (Typically Hamming Window is used)

Step-3: This step involves Fast Fourier Transform, this converts each frame of N samples from the time domain into the frequency domain.

Step-4: The scale of frequency is converted from linear to mel scale.

Step-5: Logarithm is taken from the result

Step-6: The log mel spectrum is converted back to time domain. And the final result is called the Mel Frequency Cepstrum Coefficients (MFCC).

The cepstral representation provides a good representation of the local spectral properties of the signal. Using triangular filter bank, we obtain significant decrease in the amount of data. The final results that are generated in the above process of training and test utterances are the inputs of the classification step.

3.2 Classification:

This step is very crucial in our mechanism. This includes Vector Quantization algorithm. Vector quantization is a technique which is used for audio/speech coding. The training material is used to estimate a codebook. This includes mean vectors of feature vector clusters which are given indices in order to identify them. For audio compression, the index number of nearest cluster is used instead of the original feature vector. In order to reconstruct the original audio, a revertible feature computation method has to be chosen (i.e. the above extracted MFCC features cannot be used for speech coding). The mean distance between original feature vectors and nearest mean vectors in feature space is considered as Quantization error, this also depends on the similarity between training material used for estimation of the codebook and the tested audio signal that is compressed. The classification of unknown test signals is based on the quantization error [12, 13]. For an identification decision, the error of the test feature vector sequence in regard to all codebooks is computed. If this computation error is very less, then the suspected movie with the extracted audio is considered as pirated one.

The output signal s(n) is

$$s(n) = e(n)*\theta(n) \qquad - (1).$$

Using Fourier transform we have:

$$s(w) = E(w)*\theta(w) \qquad - (2).$$

By taking logarithm we have:

$$\text{Log}|s(w)| = \log|E(w)| + \log|\theta(w)| \qquad - (3).$$

This equation is represented as follows

$$cs(w) = ce(w) + c\theta(w) \qquad - (4).$$

Using DFT the cepstral coefficients are obtained

$$cs(n) = ce(n) + c\theta(n) \qquad - (5).$$

3.2.1 Vector Quantization Process:

VQ is a process of mapping vectors of large vector space to a finite number of regions in that space.

Each region is called a cluster and is represented by its center (centroid). A collection of all centroids make a codebook. The amount of data is significantly less as the numbers of centroids are at least ten times smaller than the total number of vectors in the original sample. This reduces the number of data and the amount of computations.

Though codebooks have some spectral distortion it can still accurately represents command characteristics.

3.2.1.2 Codebook Generation:

The algorithm is implemented by the following recursive procedure

- Step-1: Design a 1-vector codebook, this is the centroid of the entire set of training vectors, this step does not require any iterations.
- Step 2: Double the size of codebook by splitting every current codebook Y_n.Where $1 \le n \le$ current code book size.
- Step-3: Using k-means iterative algorithm, nearest neighbor for each training vector is calculated.
- Step-4: Update the centroid in each cell using the centroid of the training vectors assigned.
- Step-5: repeat the steps 3 and 4 until the average distance falls below a present threshold (Iteration-1)
- Step-6: repeat steps 2, 3 and 4 until codebook of size M is reached

3.3 Recognition Phase:

In this command matching phase the features of suspected movie audio are extracted and represented by a feature vectors $\{x_1 \ldots x_n\}$. Each feature vector of sequence x is compared with already stored code words c1-cm in codebook c. For each codebook compute the distance measure and chose the lowest distance (quantization error computation). We measure the distance:

$$D=(\sum(x_i-y_j)^2)^{1/2} \qquad - (6)$$

The search of the nearest vector is done exhaustively, by finding the distance between the input vector x from test data and each of the respective code word of trained data. The one with smallest quantization error distance is considered as the output command. If this distance is very low then the fingerprints of tested and trained one are considered as similar and the suspected movie will be stated as a pirated one, else if these fingerprints doesn't match with the trained one then the movie is not a pirated copy of suspected one.

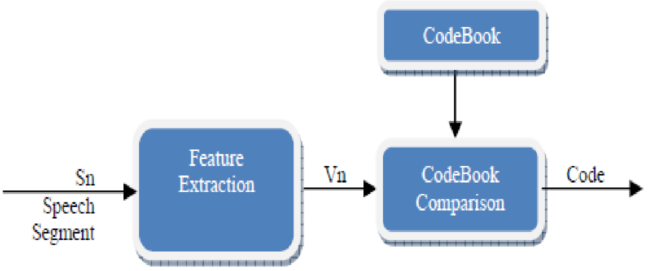


Fig. 2. Vector Quantization Encoding Process

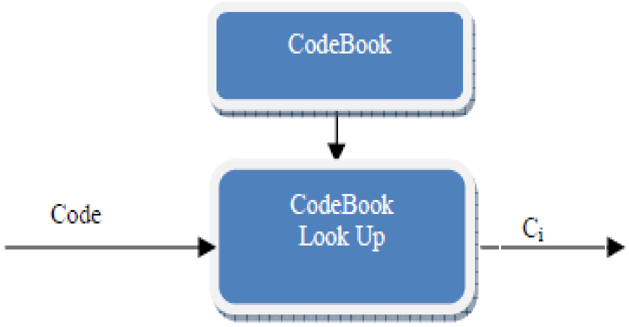


Fig.3. Vector Quantization Decoding Process

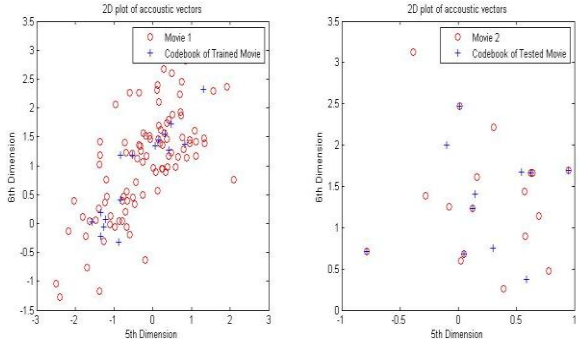
IV. RESULTS

To evaluate the efficiency of our proposed system, we registered different movies with our trained system and collected different audio samples with and without noise belonging to that movie and extracted the key features using MFCC algorithm and these key features are further classified using Vector Quantization. The resulted codebooks that are generated while implementing VQ with testing and training samples are compared for similarities. If these similarities are closer and more in number then the suspected movie will be stated as a pirated copy.

Table 1: Movie Detection using Vector Quantization with Codebook.

	Without Noise	With Noise
Training Number	52	48
Testing number	52	48
Correct Detection	50	44
Success Classification	96%	92%

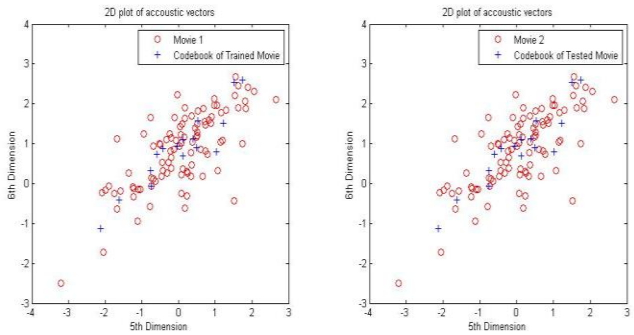
We initially took 52 samples without noise and 48 samples with noise for our experiment and tested all those samples. And compared the codebooks of trained samples over the tested ones and finally we found very less error detection rate even in the case of noised samples.



(a) Trained Sample **(b) Tested Sample**

Fig. 3 Code Books of different movies

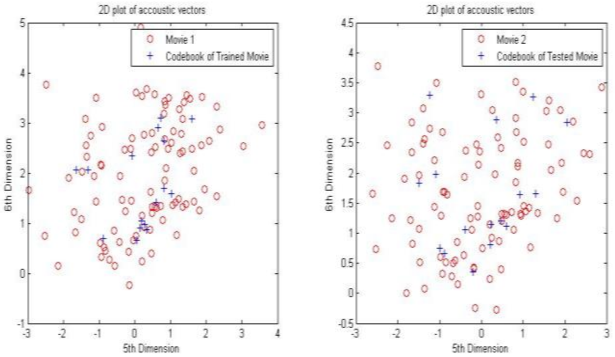
Figure 3 shows that comparison of two different movies i.e. one movie from testing movie is compared against the other movie from the training database and in the above figure we can easily find the pattern variations of codebooks generated for different movies.



(a) Trained sample **(b) Tested sample**

Fig. 4. Code Books for same movies at same Timestamps.

Figure. 4. Resembles the similarities between the codebook patterns of same movie audio samples which have same time stamps.



(a) Trained Sample (b) Tested Sample

Fig.5. Code Books of same movies with varying Timestamps

To evaluate the efficiency of our proposed technique we considered the samples of the same movie with different timestamps in order to find the pattern similarities between those timestamps. The Figure. 5. Shows the codebooks of the same movie with varying timestamps and one can observe that some patterns are still similar even with the variation of timestamps.

Table-2: Movie Recognition and related factors

	Duration	Recognition Accuracy	Computing Time (S)
Short Sample	5 Seconds	98.17%	0.6824
Long Sample	9 Seconds	97.31%	0.8231

Table-2. Epitomizes the efficiency and performance factors to detect movie piracy which are Accuracy and computing time for different samples with varying durations.

V.CONCLUSION

In this paper we proposed a novel technique to detect movie piracy using audio recognition process. We copped this idea because the audio fingerprint for each and every movie around the world is unique and our logic behind this recognition process is if any two movies have the same audio fingerprint then the movie is stated as a pirated copy for the original copyrighted movie. In this paper, our proposed system is a trained automated detection system in which the copyrighted movies which are needed to be protected against piracy must be registered initially and the audio samples are extracted from that movie. We implemented two different algorithms in which Mel-Frequency Cepstral Coefficients (MFCC) is implemented for feature extraction and Vector Quantization (VQ) is used for classification. We also considered few factors like accuracy and computation times of various audio samples with varying timestamps to measure the performance of our proposed system. We collected nearly 52 audio samples without noise and 48 samples with noise and trained and tested these samples and finally our system showed very less error detection rate.

REFERENCES

1. Satyanand Singh and Dr. E G Rajan. Article: MFCC VQ based Speaker Recognition and Its Accuracy Affecting Factors. International Journal of Computer Applications 21(6):1–6, May 2011.
2. Lindasalwa Muda, Mumtaj began and I.Elamvazuthi, Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient(MFCC) and Dynamic Time Wrapping(DTW) Techniques. Journal of Computing, Volume 2, March 2010.
3. Satyanand Singh, Dr.E.G.Rajan, Vector Quantization Approach for Speaker Recognition using MFCC and Inverted MFCC. International Journal of Computer Applications(0975-8887) March 2011
4. Michael A. Lewis, Ravi P. Ramachandran, Cochannel speaker count labeling based on the use of cepstral and pitch prediction derived features, Pattern Recognition (the journal of Pattern Recognition society) 34 499-507, 2001
5. Shi-Huang Chen and Yu-Ren Luo, Speaker verification Using MFCC and Support Vector Machine, International MultiConference of Engineers and Computer Scientists 2009 Vol 1, IMECS 2009, March 18-20, 2009, Hong Kong
6. Laura E. Boucheron, Phillip L. De Leon and Steven Sandoval, Low Bit-Rate Speech Coding through Quantization of Mel-Frequency Cepstral Coefficients. IEEE, 2011
7. Kristleifur, Herwig, Fridrik, Jonsson, Laurent, Eff^2 Videntifier: Identifying pirated movies in real-time.
8. Marc Fetscherin, Movie piracy on peer-to-peer networks- the case of KaZaA, Elsevier, Telematics and Informatics 22 (2005) 57-70
9. T. F. Quatieri, Discrete Time Speech Signal Processing. Prentice Hall, 2002.
10. Ashish Jain, Hohn Harris, Speaker identification using MFCC and HMM based techniques, university Of Florida, April 25,2004.
11. Z. Jun, S. Kwong, W. Gang, Q. Hong.: Using Mel-Frequency Cepstral Coefficients in Missing Data Technique. EURASIP Journal on Applied Signal Processing, Vol.2004, No. 3 (2004) 340-346.
12. J. C. Brown, A. Hodgins-Davis, PJO. Miller: Classification of vocalizations of killer whales using dynamic time warping. The Journal of the Acoustical Society of America, Vol. 119, (2006) EL34-EL40.
13. A.M. Youssef, T.K. Abdel-Galil, E.F. El-Saadany, M.M.A. Salama.: Disturbance classification utilizing dynamic time warping classifier. IEEE Transactions on Power Delivery, Vol. 19, No. 1, (2004) 272-278.
14. L. Lu, S. Li, and H.J. Zhang, "Content-based audio segmentation using support vector machines," Proceeding of ICME 2001, pp. 956 – 959, 2001, Tokyo, Japan.
15. Y. Linde, A. Buzo, and R. M. Gray.: „An algorithm for vector quantizer design," IEEE Trans. Commun., vol. COM-28, no. 1, pp. 84-95, 1980.

AUTHORS PROFILE

B.Srinivas received M.Tech in computer Science and engineering in 2008 from Acharya Nagarjuna University; he has two and half years of industry and four years of teaching experience. he is currently employed as a an assistant professor in CSE department, MVGR College of Engineering. He has more than eight papers in journals.

Prof. Koduganti Venkata Rao received Ph.D in Computer Science and Engineering from Andhra University, M.Tech in (Computer Science and Technology) from Andhra University and M.Sc (computer science) from Nagarjuna University, 2008, 1999, 1994 respectively. Currently Working as Head of the Department and Professor in Computer Science and Engineering and vice-principal Vignan institute Vishakapatnam.

Prof. P. Suresh Varma received Ph.D. in Computer Science and Engineering with specialization in Communication Networks from Acharya Nagarjuna University. M.Tech in (Computer Science and Technology) from Andhra University in 1998. A.M.I.E (Computer Science and Engineering from Institute of Engineers. M.Sc (Nuclear Physics from Andhra University from 1993. Currently Principal and Professor in Computer Science, Adikavi Nannaya University, Rajahmudry.