Raga Classification Based on Novel Method of Pitch Co-Occurrence

Vibhavari Rajadnya, Kalyani Joshi

Abstract**:** ***Automatic identification of raga is a growing research area and has captured significant attention from movie making industry. It is the need of time to develop efficient tools for data mining the vast audio visual data on internet. In particular, to search for a specific raga. Applications of raga search are in musicological studies, similarity based search. Ascending and descending pattern of swaras is an important feature in the raga classification. Pitch tracks of swaras are obtained from raw audio recordings. This research has utilised the pattern developed due to co-occurrence of pitches of swaras for classification. This pattern gives a concise representation of the signal which contains time and frequency information of the raga. K Nearest Neighbour (KNN) has been used as the classifier.***

Keywords**:** ***Data Mining, Music information retrieval, Raga Classification, KNN, Indian Classical Music***

I. INTRODUCTION

Music in any form is enjoyed more as compared to normal speech. There are many factors for this: like human psychology, ability to perceive, ability to respond, musicological knowledge. Melody or raga is at the heart of Indian classical music [ICM]. [1] gives the application of raga recognition related to music information retrieval [MIR] and describes in detail its significance for ICM. It also addresses the applications in music information retrieval [MIR]. In Hindustani music, a singer begins with repetitions of the raga characteristic phrases. These define the uniqueness of the raga. These characteristic phrases known as pakads, are represented by a sequence of notes in raga documentation, and there exists definite grammar rules to interpret them. Artist can take liberty of vocally expressing them without disturbing the grammar rules. Chordia [2] and Krishnaswamy [3] give appropriate definition of raga. It says, "Raga is a collection of melodic atoms and a technique for developing them. These melodic atoms are sequences of notes that are inflected with various micro pitch alterations and articulated with expressive sense of timing. Longer musical phrases are built by knitting these melodic atoms together". We can infer that raga exists as a whole and It indicates that raga is not just progression of notes. It is possible to compose every time a new tune using the same notes of raga. Various characteristics of raga on which work has been done so far include modulation or intonation of swaras, their position in the specified octave, their ascending and descending order of notes, and time spent on each note. Gamaka is an important clue for raga classification. There are about 15 types of gamakas [4, 5]. Amongst other properties is Alankar which is more of a decorative item essential to raga classification. These are patterns of note sequences which generates beautiful feelings when listening. Another most important cue for raga classification is calan. There are definite rules related to transition of notes of a raga (arohana-avarohana), the duration of each swara and rise and fall of the voice on each swara. This is defined by the calan of the raga [6]. There is an increasing trend in research related to raga classification. Main reason been tremendous amount of unclassified data on internet. Current limitations for classifying raga are because of the drawbacks of the tools and techniques for requirements of ICM [7]. In this work, we address the research of raga classification using novel technique of co-occurrence of pitch. Applications of raga identification can be identification of Gharana literally meaning 'house' which links musicians by lineage and adhere to particular music style. Applications of raga retrieval are many but not limited to information filtering systems, similarity based search, studies related to music [8], in music learning, pace tracking and speed estimation of music. Many computational opportunities spawn as a result of raga analysis [9]. Data processing of raga opens up many opportunities for Indian classical music. This paper is organised as follows. In section 2, we discuss related work for raga recognition along with various audio features used for classification. In section 3, proposed method is elaborated. Section 4 comprises of results and discussion. Section 5 covers conclusion.

Manuscript received on 25 March 2022.
Revised Manuscript received on 27 March 2022.
Manuscript published on 30 May 2022.
* Correspondence Author
Vibhavari Rajadnya*, Department of E&TC, Modern College of Engineering, Pune (Maharashtra), India. Email: vibhavari.rajadnya@gmail.com
Dr. Kalyani Joshi, Department of E&TC, Modern College of Engineering, Pune (Maharashtra), India. Email: krjpune@gmail.com

II. RELATED WORK

The most important components for any classification technique are features. Selection of audio features for raga classification is crucial as they should represent the music signal distinctly. These aids in classification in order to improve accuracy in effect to reduce the prediction error. Every research captures features from a particular perspective. In [10], [29] authors have made use of 'pakad' as feature. Pakad matching was performed using N gram matching. Pitch as a feature is used and 'Praat' software is used. 'Chromogram' of ragas is used by authors of [11], [12] and classifier used is Bayesian classifier. [13], [14], [15], [16] make use of Pitch Class Distribution (PCD) features.

It is based on tonal expectation and perception. Weighted pitch class distributions is used to form a tonal hierarchy for minor ad major keys. Overall frequency usage in the song is well captured through PCD. PCDs have proven to be powerful against pitch-octave errors. PCD-based approach is currently the most appropriate and accurate method and by far the most researched feature for raga recognition [6]. Another paper based on PCD is in [17], Chordia et al. used dataset which consisted of 17 ragas which were monophonic sung by a single artist using the PCD based approach. Pitch as feature is extracted using the Harmonic Product Spectrum (HPS) algorithm. In the audio clip, presence of note is found by sensing the abrupt changes in amplitude and phase of the signal spectrum. After this, calculation of pitch-class profiles and the bigrams is completed. It is proved that allied ragas are distinguished accurately using bi-grams. However PCDs have short coming that they do not carry sequential information which can prove to be useful in distinguishing allied ragas [18]. In particular for raga classification problem, along with frequency, temporal or sequential features are also equally important [19]. These are particularly useful for distinguishing allied ragas which have same set of svaras. One of the technique of capturing temporal characteristics involves constructing a representation for the melody. Authors of [20, 21, and 22] tried this using n-gram distributions [20], hidden Markov models [21, and 28], and melodic progression templates [22]. The sequential information is extracted using the arohana-avrohana pattern of a raga. Shortcoming of this method is that they don't take into consideration the peculiar melodic transitions in between svaras called as gamaks, which have shown to be useful in discriminating ragas [23]. Based on the above findings, in our research, we propose to use pitch as feature. The reasons for this choice are: firstly pitch based PCD approach has shown promising result [6, 17]. As a result it is rated high when it comes to feature selection for raga recognition problem and secondly authentic, diverse, largest database of Carnatic which consists of predominant pitch required for the research is publicly available. [30] gives useful information by performing statistical comparisons of two ragas Bhairav and Bihag. In this paper, co-occurrence matrix is used to capture the pattern of pitch in different frames. Second order statistics [SOS] are computed for the co-occurrence matrix. These are correlation, energy, homogeneity, contrast, entropy are computed. These measures represent a pattern of pitch and used as features for classification. In order to verify for usefulness of second order statistical parameters, it was decided to capture the pattern of these parameters across multiple co-occurrence matrices. These matrices were calculated by introducing a delay in between frames. This delay was more related to stationary duration of the signal. The delay parameter was varied from 0.1s to 1.5s in step of 0.1s and co-occurrence matrix along with 5 parameters were calculated for each of these. Thus with this approach temporal as well as sequential features are captured effectively. For classification, we use k-Nearest neighbour classifier and three standard distance measures. A simple algorithm to implement especially when it has to be for large number of classes. Execution speed is high with minimal training. Its main characteristics is its single parameter k to be required to be tuned.

III. PROPOSED APPROACH

A. Computation of co-Occurrence Matrix

Matlab is used as the development platform. For computation of co-occurrence matrix we referred steps mentioned in [31]. With this a 120 by 120 co-occurrence matrix was obtained for each recording. It describes how frequently two bin pairs in 2 successive frames appear in the window separated by a distance d in direction θ. This matrix efficiently captures both time and frequency information of different svaras. It computes how frequently two pitch pairs appear in two consecutive frames which are some distance apart.

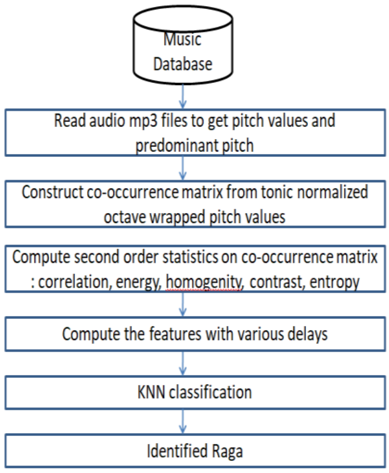

Figure III.1 Raga Identification Steps

B. Computation of Statistical Measures

Second order statistics of the co-occurrence matrix namely homogeneity, energy, correlation, contrast entropy are computed. It clearly shows the pattern of the predominant pitch across entire audio clip.

Correlation

Measure of how correlated a bin value (which represents normalised pitch in this project) in one frame is to its neighbouring frame in a mp3 file. If the frames are positively correlated it will be 1 else -1. Can be used to assess the regularity in pitch in the mp3 file.

$$Correlation = \sum_{i,j} \frac{(X_i-\bar{X}_i)(Y_j-\bar{Y}_j)}{\sum(X_i-\bar{X}_i)^2 \sum(Y_j-\bar{Y}_j)^2} \quad (3)$$

Where:
X is the co-occurrence matrix
- mean of the matrix in row direction
- mean of the matrix in column direction

Energy

It calculates addition of squared elements. It is an indication of how similar are the pitch vaue levels in adjacent frames. If the pitch is constant, it will return 1.

$$Energy = \frac{1}{n}\sum_{m=0}^{n-1}[X(m)] \quad (4)$$

Entropy

It is an indication of variation in pitch values in frames. If the pitch is constant, all thru the files, it will be 0.

Retrieval Number: 100.1/ijrte.A68860511122
DOI: 10.35940/ijrte.A6886.0511122
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

$$Entropy = -\sum_{i=1}^{n} X_i \log X_i \qquad (5)$$

Homogeneity

Homogeneity is a measure which indicates how close the elements in the matrix are with reference to its diagonal. If the pitch values in the most of the have similar values then they will be near to the diagonal. If the differnce in pitch values in successive frames is 10 cents thne they will one cell away (1-2, 2-3 etc.) from the diagonal. If the difference in pitch values is greater then they will be far away from the diagonal.

$$Homogeneity = \sum_{i,j} \frac{X(i,j)}{1+|i-j|} \qquad (6)$$

Contrast

It computes the difference in pitch values between two neighbouring frames in the entire recording file. If it is constant value, contrast will be 0

$$Contrast = \sum_{i,j} |i-j|^2 \; X(i,j) \qquad (7)$$

C. Computing Pattern of Statistical Measures by varying delay in two frames

In order to verify for usefulness of second order statistical parameters, it was decided to capture the pattern of these parameters across multiple co-occurrence matrices. These matrices were calculated by introducing a delay in between frames. This delay was more related to stationary duration of the signal. The delay parameter was varied from 0.1s to 1.5s in step of 0.1s and co-occurrence matrix along with 5 parameters were calculated for each of these. The patterns are self explanatory for their usefulness in raga classification.
Feature vector is constructed using these 5 statistical measures at various delays and the 120 by 120 co-occurrence matrix. This feature vector is used for classification

D. Classification and distance measurement

Music files in the database are classified according to raga labels. kNN is used as classifier [24], [25], [26]. It is very simple algorithm to understand and interpret. KNN uses two parameters i.e., distance function (e.g. Euclidean or Manhattan etc.) and the value of K and zero training period. Algorithm does not learn any hyper plane distinguishing function from the training data. Due to absence of training, new rcordings can be added easily as this will not influence the algorithm's accuracy.

E. Implementation steps:

We have used 3 different distance measures, to calculate distance between 2 recordings. S(n) and S(m) are feature vectors of 2 consecutive recordings in the below formulaes:

- **Frobenius norm:**

$$D_F^{(n,m)} = \|S_n - S_m\|_2 \qquad (8)$$

Where Sn = nth recording's feature vector
Where Sm = mth recording's feature vector

- **Kullback-Leibler divergence:**

$$D_{KL}^{(n,m)} = D_{KL}\left(S^{(n)}, S^{(m)}\right) + D_{KL}\left(S^{(m)}, S^{(m)}\right)$$

$$D_{KL}(X,Y) = \sum X \log\left(\frac{X}{Y}\right)$$

- **Bhattacharya:**

$$D_B^{(n,m)} = -\log\left(\sum \sqrt{S^{(n)}.S^{(m)}}\right)$$

For all the three distances, each element is considered for operations and then addition is performed over all the elements of the feature vector. Three values of k were taken for experimentation: 1, 3, and 5.

F. Testing

We have made use of database of 480 recordings of 40 ragas from Carnatic music dataset (CMD) consisting of 124 hours of recordings. They are stored as 160kbps mp3 files [6], [27]. It is composed of 12 recordings for each class. Thus the dataset is balanced in the number of recordings per class. Leave one out cross validation strategy is used for testing. In involves all but one recording from the database used for testing and rest all form the training set. Majority class from the k nearest neighbours of the classified ragas is assigned to the unknown recording. If the majority classes are equal in number, selection is done arbitrarily

IV. RESULTS AND DISCUSSION

As can be inferred from the results, raga classification task becomes hard as the number of classes increase. Figure 2 shows pattern of SOS parameters for various delays for CMD dataset. Pattern for contrast and entropy features can be seen to contribute highest for raga classification. Entropy can be attributed to the high difference in variation in pitch values at various delays. Contrast measures the difference in pitch values which is seen to be high for specific delays. This is because of different set of svaras used in the avarohana and arohana of the raga. These patterns of co-occurrence matrix SOS features, also effectively captures the melodic transitions between svaras.

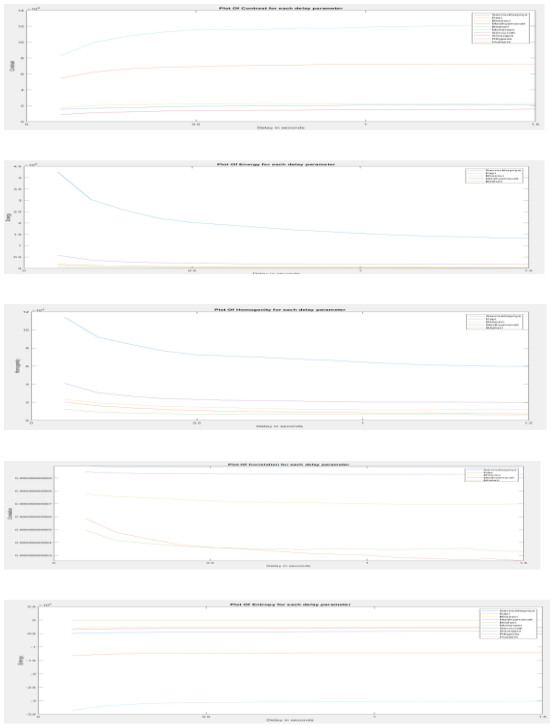

Figure IV.1 Plot of pattern of contrast, energy, homogenity, correlation, entropy for co-occurrence matrix calculated for various delays of CMD for raga Madhyamavati, Sanmukhapriya, Bhairavi, Bilahari, Kapi from CMD respectively

Confusion matrix displays the relationship between predicted and actual classification done by the classifier. For accuracy purpose, it is calculated over the predicted classes. Figure 3 shows the CMD confusion matrix plot. The accuracy is 78.1% for Bhattacharya distance.

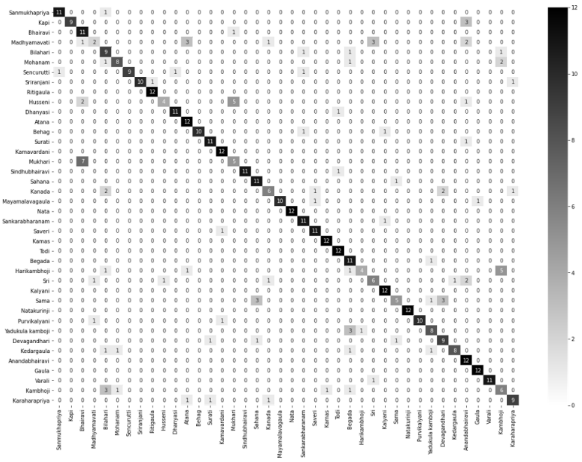

Figure IV.2: CMD Confusion Matrix with Bhattacharya distance

Accuracy for k=1, 3, 5 is captured in fig 4 with 3 distance measures. It is 78.1% for k=3 for CMD which is highest for Bhattacharya distance. It is observed that accuracy decreases with the increase in number of nearest neighbours. A better result can be obtained if number of recordings per class are increased.

K	Frobenius (%)	KLD (%)	Bhattacharya(%)
1	71	77.5	78.6
3	69.6	76.7	78.1
5	70.6	76.5	77.9

Figure IV.3 Accuracy plotted as a function of 'number of nearest neighbours' for CMD

For verification purpose, we use two existing methods for evaluation: Gulati [27], denoted by EVSM, Chordia & Senturk [13], denoted by EPCD. Figure 5 captures for the CMD dataset, results for the existing methods and the current method for three different distances. Features based on Pitch Class Dyad are computed from the entire audio recording for EPCD where as automatically discovered melodic phrases and vector space modelling is used in EVSM. Peak accuracy of 78.1% on CMD is obtained for Bhattacharya distance. It is higher than 73.1% obtained by EPCD and 68.1% by EVSM for CMD.

Data Set	Frobenius (%)	KLD (%)	Bhattacharya (%)	EPCD (%)	EVSM (%)
HMD	85.3	93	93.3	91.7	83
CMD	69.6	76.7	78.1	73.1	68.1

Figure IV.4 Accuracy of EPCD and EVSM as the two existing methods and three proposed variants.

V. CONCLUSIONS

Till date there are many techniques using varied features for classification. Each feature has a specific purpose based on raga characteristics. Pitch co-occurrence pattern technique is giving accuracy of 78.1% for CMD dataset. Tonal and spatial characteristics of ragas for their classification are successfully giving better acuracy. It is particularly helpful in distinguishing allied ragas. To conclude, use of Deep Neural Network (DNN) can be explored to learn the temporal and tonal patterns.

REFERENCES

1. G. Geekie, "Carnatic ragas as music information retrieval entities," in Proc. of ISMIR, 2002, pp. 257–2
2. P. Chordia and A. Rae, "Raag recognition using pitchclass and pitch-class dyad distributions," in Proc. of ISMIR, 2007, pp. 431–436.
3. Krishnaswamy, "Melodic atoms for transcribing carnatic music," in Proc. of ISMIR, 2004, pp. 345–348.
4. S. R. Janakiraman, Essentials of Musicology in South Indian Music. The Indian Music Publishing House, 2008.
5. P. P. Narayanaswami and V. Jayaraman, 2004. [Online]. Available: http://ibiblio.org/guruguha/ssp.htm
6. S. Gulati, J. Serrà, K. K. Ganguli, S. Şentürk, and X. Serra. Time-delayed melody surfaces for rāga recognition. In Int. Soc. for Music Information Retrieval Conf. (ISMIR), pages 751–757, New York, USA, 2016.
7. X. Serra. A multicultural approach to music information research. In Proc. of Int. Conf. on Music Information Retrieval (ISMIR), pages 151–156, 2011.
8. Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang, "A Survey of Audio-Based Music Classification and Annotation", IEEE Transactions on Multimedia, Vol. 13, No. 2, April 2011.
9. "An Overview of Hindustani Music in the Context of Computational Musicology" Journal of New Music Research, Volume 43, Issue 1, 31 Mar 2014 by Suvarnalata Rao and Preeti Rao.
10. H. Sharma, R. S.Bali , "Comparison of ML classifiers for Raga recognition", International Journal of Scientific and Research Publications, Volume 5, Issue 10, October 2015, ISSN 2250-3153
11. G. K. Koduri, V. Ishwar, J. Serra, and X. Serra, "Into-nation analysis of ragas in Carnatic music", Journal of New Music Research, 43(1):72–93, 2014.
12. T. M. Krishna and V. Ishwar. Karnatic music: Svara, gamaka, motif and raga identity. In Proc. of the 2nd CompMusic Workshop, pages 12–18, 2012.
13. K. C. Waghmare and B. A. Sonkamble, "Analyzing Acoustics of Indian Music Audio Signal Using Timbre and Pitch Features for Raga Identification," 2019 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC), 2019, pp. 42-46, doi: 10.1109/ICISPC.2019.8935707.
14. P. Dighe, P. Agrawal, H. Karnick, S. Thota, and B. Raj. Scale independent raga identification using chromagram patterns and swara based features. In IEEE Int. Conf. on Multimedia and Expo Workshops (ICMEW), pages 1–4, 2013.
15. P. Dighe, H. Karnick, and B. Raj. Swara histogram based structural analysis and identification of Indian classical ragas. In Proc. of Int. Soc. for Music Information Retrieval Conf. (ISMIR), pages 35–40, 2013.
16. G. K. Koduri, V. Ishwar, J. Serra, and X. Serra. Intonation analysis of ragas in Carnatic music. Journal of New Music Research, 43(1):72–93, 2014
17. P. Chordia and S. Senturk. Joint recognition of raag and ¨ tonic in north Indian music. Computer Music Journal, 37(3):82–98, 2013.
18. S. Rao, J. Bor, W. vander Meer, and J. Harvey. The raga guide: a survey of 74 Hindustani ragas. Nimbus Records with Rotterdam Conservatory of Music, 1999.
19. T. M. Krishna and V. Ishwar. Karnatic music: Svara, gamaka, motif and raga identity. InProc. of the 2nd CompMusic Workshop, pages 12–18, 2012.
20. V. Kumar, H Pandya, and C. V. Jawahar. Identifying ragas in Indian music. In 22nd Int. Conf. on Pattern Recognition (ICPR), pages 767–772, 2014.
21. P. V. Rajkumar, K. P. Saishankar, and M. John. Identification of Carnatic raagas using hidden markov models. In IEEE 9th Int. Symposium on Applied Machine Intelligence and Informatics (SAMI), pages 107–110, 2011.
22. S. Shetty and K. K. Achary. Raga mining of Indian music by extracting arohana-avarohana pattern. Int. Journal of Recent Trends in Engineering, 1(1):362–366, 2009.
23. S. Rao, J. Bor, W. vander Meer, and J. Harvey. The raga guide: a survey of 74 Hindustani ragas. Nimbus Records with Rotterdam Conservatory of Music, 1999.

Retrieval Number: 100.1/ijrte.A68860511122
DOI: 10.35940/ijrte.A6886.0511122
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

24. T. M. Mitchell. Machine Learning. McGraw-Hill, New York, USA, 1997.
25. Guo, G., Ping, X., & Chen, G. (2006). A Fast Document Classification algorithm based on improved KNN, 3–6.
26. Hassanat, A. B., Abbadi, M. A., & Alhasanat, A. A. (2014). Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach. International Journal of Computer Science and Information Security (IJCSIS), 12(8), 33–39. doi:10.1007/s00500-005-0503-y
27. S. Gulati, J. Serrà, V. Ishwar, S. Şentürk, and X. Serra. Phrase-based rāga recognition using vector space modeling. In IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), pages 66–70, Shanghai, China, 2016
28. Shetty S., Achary K.K., Hegde S. (2012) Raga Identification in CARNATIC Music Using Hidden Markov Model Technique.
29. Tewari S., Chakraborty S. (2020) Modeling a Raga-Based Song and Evaluating Its Raga Content: Why It Matters in a Clinical Setting. In: Sahana S., Bhattacharjee V. (eds) Advances in Computational Intelligence. Advances in Intelligent Systems and Computing, vol 988. Springer, Singapore.
30. Chakraborty S., Mazzola G., Tewari S., Patra M. (2014) A Statistical Comparison of Bhairav (a Morning Raga) and Bihag (a Night Raga). In: Computational Musicology in Hindustani Music. Computational Music Science. Springer, Cham.
31. Vibhavari Rajadnya, Kalyani Joshi, "Indonesian Journal of Electrical Engineering and Computer Science" Vol. 24, No. 1, October 2021, pp. 158~167 ISSN: 2502-4752, DOI: 10.11591/ijeecs.v24.i1.pp158-167

AUTHORS PROFILE

Vibhavari Rajadnya, has completed her bachelor's degree in Electronics amd Telecommunications from Cummins College of Engineering, Pune, Maharashtra, India in year 1996. She completed her Masters degree in Electronics Engineering from Cummins College of Engineering, Pune, Maharashtra in year 2013. Currently she is pursuing her Ph. D. in Electronics and Telecommunications engineering from Modern College of Engineering, Pune under Pune University. Having an experience of 14 years in the IT industry, she is currently working as a project manager with Fairshare IT services Ltd, Pune, Maharashtra, India. Her research interests include music signal Image processing.

Dr. Mrs. Kalyani R. Joshi, has currently been working as Principal and Professor E&TC in PES's Modern College of Engineering, Pune, Maharashtra, India. She has completed her Bachelor's degree in Electronics Engineering from Shivaji University and Masters degree in E&TC from College of Engineering, Pune under Pune University. She has got her Ph. D. degree from College of Engineering, Pune in year 2008 with specialization in biomedical image processing. She has been worked as Chair, Signal Processing Chapter, IEEE Pune section for three consecutive years from 2013 to 2015 and awarded as senior member of IEEE. She has Coauthored a book on Image And Video Compression Fundamentals, CRC Press-2015. She has more than 30 national and international publications in reputed journals and conferences. Her areas of interests are Digital Signal and Image Processing and signal processing in Communication.

Retrieval Number: 100.1/ijrte.A68860511122
DOI: 10.35940/ijrte.A6886.0511122
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)