

Text Localization in Video Data Using Discrete Wavelet Transform

G.Nagendhar¹, D.Rajani², China Venkateswarlu Sonagiri³, V.Sridhar⁴

Asst.Prof-ECE VJIT- JNTUH¹, Assoc.Prof-ECE VJIT- JNTUH², Professor& HOD-ECE HITS -JNTUH³

Asst. Prof-ECE VJIT- JNTUH⁴

Abstract: Text provides important information about images or video sequences in a documented image, but it always remains difficult to modify the static documented image. To carry out modification in any of the text matter the text must be segmented out from the documented image, which can be used for further analysis. Taking consideration to video image sequence the isolation of text data from the isolated frame becomes more difficult due to its variable nature. Various methods were proposed for the isolation of text data from the documented image. Among which Wavelet transforms have been widely used as effective tool in text segmentation. Document images usually contain three types of texture information. various wavelet transformation have been proposed for the decomposition of these images into their fundamentals feature. Onto these wavelet families, it is one of the difficult tasks in selecting a proper wavelet transformation with proper scale level for text isolation. This paper work implements an efficient text isolation algorithm for the extraction of text data from the documented video clips. The implemented system carries out a performance analysis on various wavelet transforms for the proper selection of wavelet transform with multi level decomposition. Of the selected wavelet transform the obtained wavelet a coefficient are applied with morphological operators for text isolation and evaluates the contribution of decomposition levels and wavelet functions to the segmentation result in documented video image. The proposed task implements neural network for the recognition of text characters from the isolated text image for making it.

Keywords: decomposition, *images, isolation Wavelet transforms, Segmentation,*

I. INTRODUCTION

Traditional medium for printed documents. However, with the advancement of digital technology, it is seen that paper documents gradually augmented by electronic documents. Paper documents consist of printed information on paper media. Electronic documents use predefined digital formats where information regarding both textual and graphical document elements, have been recorded along with layout and stylistic data. Both paper and Electronic documents confer their own advantages and disadvantages to the user. For example, information on paper is easy to access but tedious under modification and difficult under storage of huge information's, while Electronic documents are best under storage of huge data base but very difficult under modifications.

In order to gain the benefits of both media, the user needs to be able to port information freely between the two formats. Due to this need, the development of computer systems capable of accomplishing this interconversion is needed. Therefore, Automatic Document Conversion has become increasingly important in many areas of academia, business and industry. Automatic Document Conversion, occurs in two directions: Document Formatting, and Document Image Analysis The first automatically converts Electronic documents to paper documents, and the second, converts paper documents to their electronic counterparts. Document Image Analysis is concerned with the problem of transferring the document images into electronic format. This would involve the automatic interpretation of text images in a printed document, such as books, reference papers, newspapers etc. Document Image Analysis can be defined as the process that performs the overall interpretation of document images. It is a key area of research for various applications in machine vision and media processing, including page readers, content-based document retrieval, digital libraries etc. There is a considerable amount of text occurring in video that is a useful source of information, which can be used to improve the indexing of video. The presence of text in a scene, to some extent, naturally describes its content. If this text information can be harnessed, it can be used along with the temporal segmentation methods to provide a much truer form of content-based access to the video data.

II. IMAGE SEGMENTATION

Many efforts have been done to address the problems of text area detection, text segmentation, and text recognition. Current text detection approaches can be classified into three categories. The first category is connected component-based method, which can locate text quickly but have difficulties when text is embedded in complex background or touches other graphical objects. The second category is texture-based, which is hard to find accurate boundaries of text areas and usually yields many false alarms in “text-like” background texture areas. The third category is edge-based method. Generally, analyzing the projection profiles of edge intensity maps can decompose text regions and can efficiently predict the text data from a given video image clip. Text region usually have a special texture because they consist of identical character components. These components contrast the background and have a periodic horizontal intensity variation due to the horizontal alignment of many characters. As a result, text regions can be segmented using texture feature. Document Image Segmentation is the act of partitioning a document image into separated regions. These regions should ideally correspond to the image entities such as text blocks and graphical images, which are present in the document image. These entities can then be identified and processed as required by the subsequent steps of Automated Document Conversion.

Various methods are described for processing Document Image Segmentation. They include: Layout Analysis, Geometric Structure Detection/Analysis, Document Analysis, Document Page Decomposition, Layout Segmentation, etc. Texts in images and video sequences provide highly condensed information about the contents of the images or videos sequences and can be used for video browsing/retrieval in a large image database. Although texts provide important information about images or video sequences, it is not easy to detect and segment out the text data from the documented image. The difficulty in text extraction is due to the following reasons. The text properties vary randomly with non uniform distribution. Texts present in an image or a video sequence may have different cluttered background. Methods for texts extraction can be done using component-based or texture-based. Using component-based texts extraction methods, text regions are detected by analyzing the edge component of the candidate regions or homogenous color/grayscale components that contain the characters. Whereas texture based method uses the texture property such as curviness of the character and image for text isolation. In texture based document image analysis an M-band wavelet transformation is used which decomposes the image into $M \times M$ band pass sub channels so as to detect the text regions easily from the documented image. The intensity of the candidate text edges is used to recognize the real text regions in an M-sub band image.

III. IMAGE ANALYSIS

Digital image is represented as a two-dimensional array of coefficients, each coefficient representing the intensity level at that coordinate. Most natural images have smooth color variations, with the fine details being represented as sharp edges in between the smooth variations. Technically, the smooth variations in color can be termed as low frequency variations, and the sharp variations as high frequency variations. Separating the smooth variations and details of the image can be performed in many ways. One way is the decomposition of the image using the discrete wavelet transform. Digital image compression is based on the ideas of sub-band decomposition or discrete wavelet transforms. Wavelets, which refer to a set of basis functions, are defined recursively from a set of scaling coefficients and scaling functions. Image Segmentation is a crucial step in the conversion process for paper document images into electronic documents. Entities in a document image, such as text blocks, and figures need to be separated before further document analysis and recognition can occur. Many Document Segmentation algorithms are designed exclusively for a few specific document types, utilizing highly specialized document models.

Basically an independent segmenter does not assume specific document layout models in its segmentation. The segmenter utilizes a minimal amount of image domain knowledge. Entities from the document images are extracted as non-overlapping sub-images by the segmenter.

The advantages of document image analysis are:

1. Document size.

An ASCII representation of a document page can easily be stored in 2-3 KB, whereas a typical scanned image of a page may require between 500 KB and 2 MB. If it is to maintain documents in image form, an efficient compressed representation is essential for both storage and transmission.

2. Providing efficient access to the compressed image.

Traditional compression techniques used for document images have been successful in reducing storage requirements but do not provide efficient access to the compressed data. It is desirable to use a compression method that makes use of a structured representation of the data, so that it not only allows for rapid transmission but also allows access to various document components and facilitates processing of documents without the need for expensive decompression.

3. Readability. Many lossy compression and progressive transmission techniques use resolution reduction or texture-preserving methods that can render a document image unreadable. It is desirable that a document be readable even at the highest levels of lossy compression and at the start of a progressive transmission. The highly lossy representation can then be augmented by subsequent transmissions for better rendition.

The following is rough summary of the document image analysis process:

1. A paper document is scanned to a digital form as digital document image.
2. Pre processing filters are applied to reduce noise or image distortion. Binarization of the image is performed during this stage of processing.

3. Segmentation is performed on a digital document image by splitting the document image into discrete graphic entries.
4. Segmented entries are classified into different entity types.
5. Text entities are isolated, and individual characters or words in the entities are extracted.
6. Non text entities are analyzed and any text matter in these entries are also extracted.
7. Extracted text characters from the entities undergo further processing.
8. Recognized text from both text and non text entries are analyzed to extract their logical relationships with other entries in the document.

IV. VIDEO DOCUMENT TEXT EXTRACTION

In the age of multimedia, video is increasingly important and common information medium. However, most current video data is unstructured, i.e. stored and displayed only as pixels. There is no additional content information such as year of production, starring actors, director, producer, costume designer, places of shots, positions and types of scene breaks etc. Therefore, the usability of raw video is limited, precluding effective and efficient retrieval. Consider the thousand of MPEG-encoded films on the Internet. Beyond the title and a short description rarely can any information be found about the content and structure of these films, therefore making it very difficult to find e.g. specific kinds of films or scenes. Information on video content would be highly desirable.

V. DISCRETE WAVELET TRANSFORM

The discrete wavelet transform is a very useful tool for signal analysis and image processing, especially in multi-resolution representation. In image processing, it is difficult to analyze the information about an image directly from the gray-level intensity of image pixels. The multi-resolution representation can provide a simple method for expounding the information about images. The two-dimensional discrete wavelet transform can decompose an image into 4 different resolutions of sub-bands. Those sub-bands include one average sub-bands and three detail component sub-bands. Detail component sub-bands represent different features for an image. In this subsection, we introduce the basic theory of discrete wavelet transform and the 9-7 taps discrete wavelet transform filters used in image processing.

BIORTHOGONAL SPLINE WAVELETS.

Images are generally smooth. So, using smooth mother wavelet for image analysis is a good practice. On the other hand, it is also desirable that the mother wavelet is symmetric so that the corresponding wavelet transform can be implemented using mirror boundary conditions that reduce boundary artifacts. Except for the Haar wavelet, no wavelets are both orthogonal and symmetric. To achieve the symmetric property, we can relax the orthogonality requirement by using a biorthogonal basis. The following figure shows the relation between the filter structure and wavelets functions.

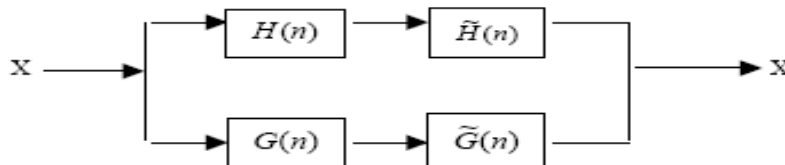


Figure V.1: Filter Structure and the Associating Wavelets

VI. INTERPRETATION OF RESULTS

The initial GUI created which shows the project title, and other details regarding the project work. In this window a continue button is used to go for further processing and a close button to terminate the application.

READING DOCUMENTED VIDEO FILE

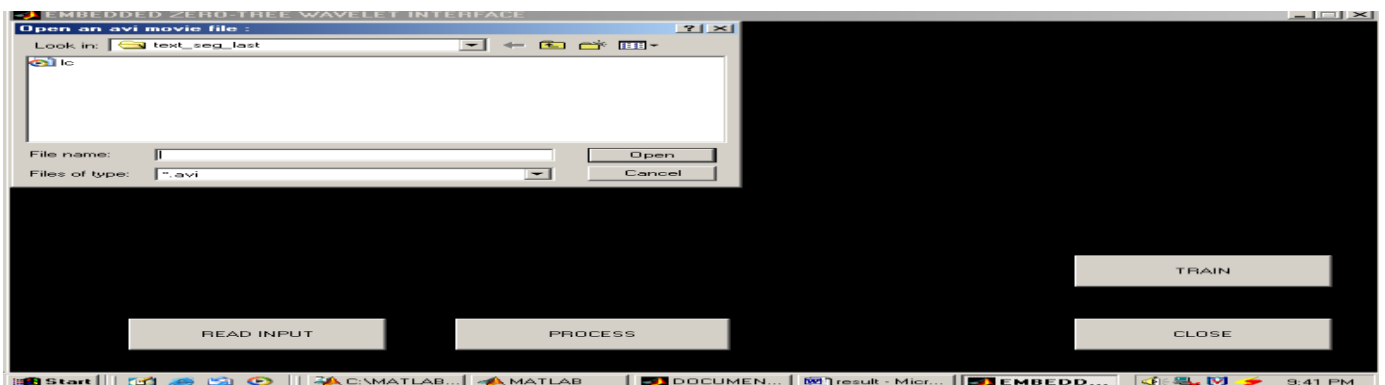


Figure VI.1: Reads documented video file

The next GUI created is for input interface having a READ INPUT button, which is used to select a desired video file for processing from the select window. The process button helps to move further in the application and close button terminate the current process.

DOCUMENTED VIDEO FILE



Figure VI.2: Documented video file

On selection of READ INPUT button the selected input get displayed for processing. On selection of TRAIN the alphanumeric data which is stored in database is trained for character recognition.

PROCESS WINDOW

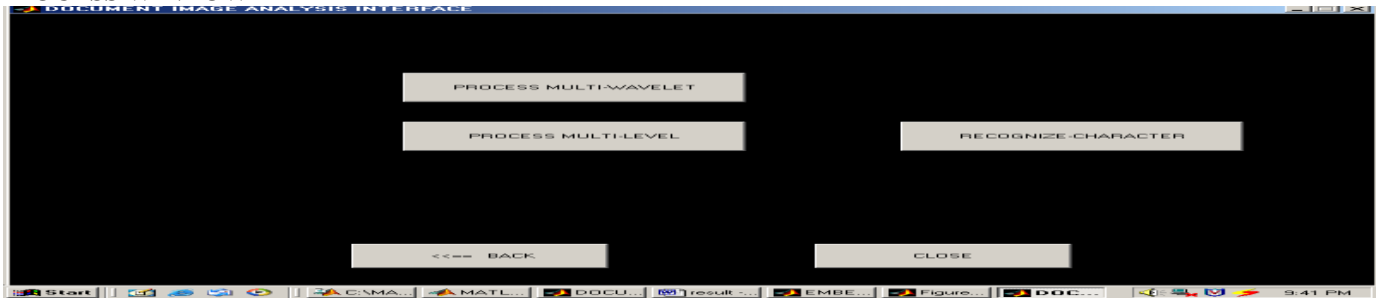


Figure VI.3:: Processing Multilevel, Multi Wavelet and Recognize Character

On selecting process multi-wavelet button the input data is decomposed into horizontal diagonal, vertical and approximate data. On selecting of process multi-level button the input data is decomposed into various levels. On selecting recognize character button the extracted text is presented in editable form.

MULTI-WAVELET TRANSFORMATION



Figure VI.4: processing documented video data by wavelets.

Documented video data is processed using various wavelets such as Haar wavelet, Debuchie wavelet and Spline wavelet.

PROCESSING INPUT DATA USING HAAR WAVELET

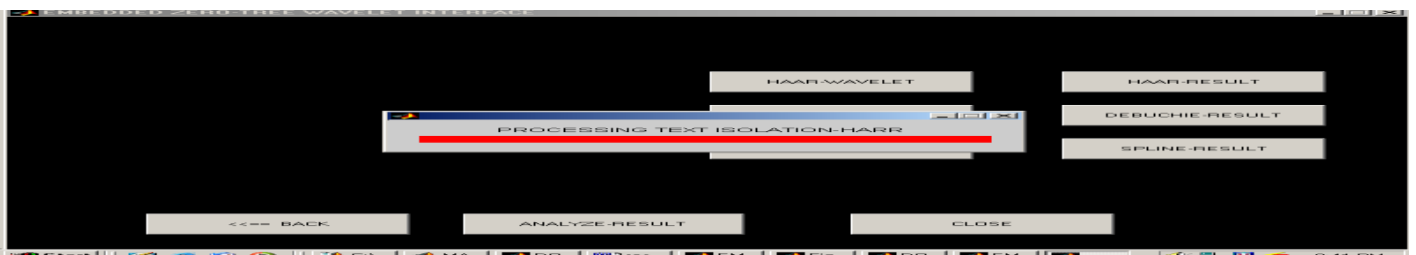


Figure VI.5: Processing video data using Haar wavelet

Input documented video data file is processed using Haar wavelet transformation to isolate the text. Similarly, input data is processed for Debuchie and spline wavelets.

TEXT ISOLATION USING HAAR WAVELET

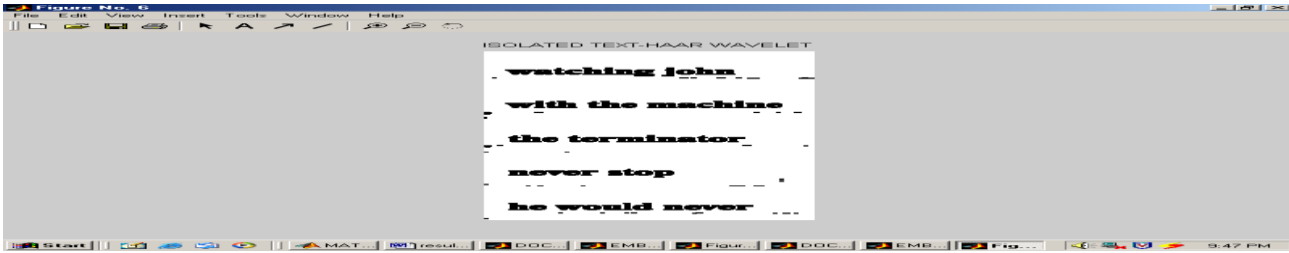


Figure VI.6: text isolation from video file using Haar wavelet

Figure shows the text, which is isolated from documented video data file using Haar wavelet transformation.

ERROR COMPARISON PLOT FOR WAVELETS

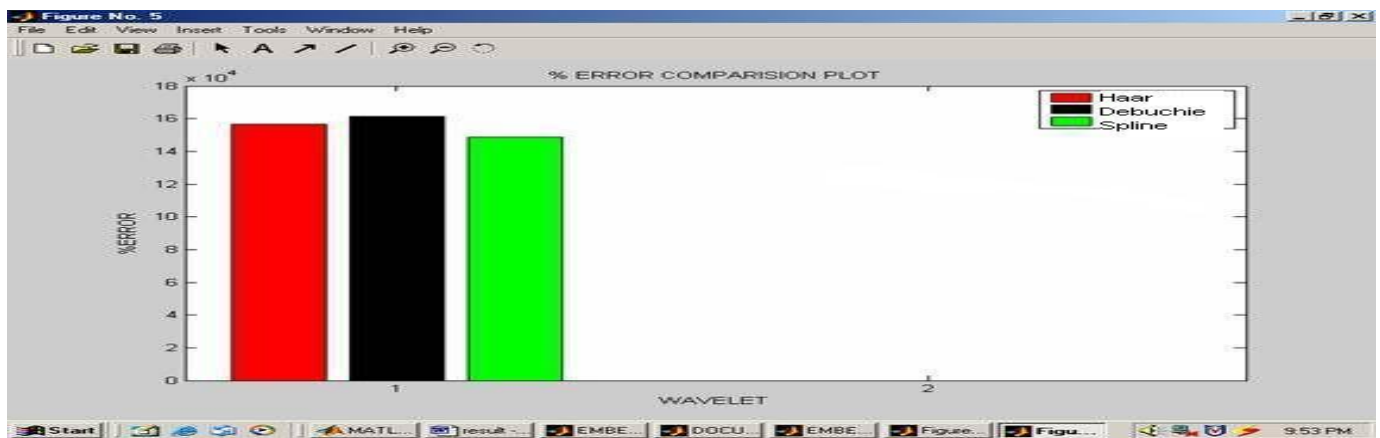


Figure VI.7: Error comparison plot for WAVELETS

The above figure analysis the % of error for multi level wavelets. It is observed that percentage of error is less in spline wavelet. . The analysis plots were obtained on selection of analyze result button on the GUI.

TIME COMPARISON PLOT FOR WAVELETS

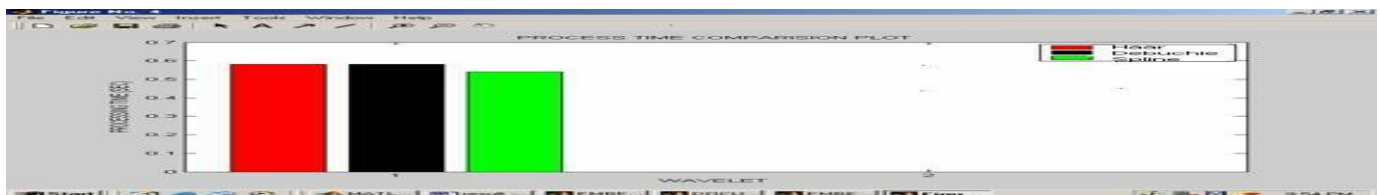


Figure VI.8: Processing time comparison plot for WAVELETS

The work analyzes the efficient wavelet by comparing the time of processing taken by the Haar, Debuchie and Spline wavelets. The analysis plots were obtained on selection of analyze result button on the GUI.

LEVEL 1 SCALED IMAGE



Figure VI.9: level 1 scaled image

The work analyses the efficient multi level scaling by comparing the time of processing taken by the level 1 scaling, level 2 scaling, level 3 scaling and level 4 scaling. The analysis plots were obtained on selection of analyze result button on the GUI.

Table No. VI.1: Comparison table for wavelets

| Wave let | Processing time (sec) | % Error |
|----------|-----------------------|---------|
| Haar | 0.58 | 15.7 |
| Debuchie | 0.58 | 16.2 |
| Spline | 0.55 | 15 |

VII. CONCLUSION

The paper aims towards the realization of text segmentation algorithm for the isolation text data in a documented image. The project also aims towards the testing of spline wavelet compatibility for text isolation in a documented image over other wavelet transformations. The segmentation systems implemented uses Haar, Daubechie-- wavelets and compare the obtained results with the results obtained for spline wavelets. The segmentation system is tested for different properties of images. The first sample considered is taken as a uniformly distributed text and graphic where as the second sample is non-uniform in nature. The 3rd sample is considered to be overlapped documented image where the text data is present over the image. For the entire three considered documented image an error analysis is carried out considering different wavelet transform and at different levels. From the result obtained it is clearly observed that biorthogonal Spline wavelet gives more accurate segmentation over other wavelet transformations. The present proposed work is carried out on one layered or 2 layered documented images. The work can be extended for still higher layered documented image. Work can be carried out on consideration to other parameters such as documented image intensity.

REFERENCES

[1] Chung-Wei Liang and Po-Yueh Chen, "DWT Based Text Localization", Int. J. Appl. Sci. Eng., 2004. 2, 1.
 [2] Jie Xi, Xian-Sheng Hua, Xiang-Rong Chen, Liu Wenyin, Hong-Jiang Zhang, "A Video Text Detection And Recognition System", Microsoft Research China 49 Zhichun Road, Beijing 100080, China.
 [3] Xian-Sheng Hua, Pei Yin, Hong-Jiang Zhang. "Efficient Video Text Recognition Using Multiple Frame Integration", Microsoft Research Asia, 2.
 [4] C'eline Thillou and Bernard Gosselin, "Robust Thresholding Based On Wavelets And Thinning Algorithms For Degraded Camera Images", Facult'e Polytechnique de Mons, Avenue Copernic, 7000 Mons, Belgium.
 [5] Celine Thillou, Bernard Gosselin, "Segmentation-Based Binarization For Colordegraded Images ", Facult'e Polytechnique de Mons, Avenue Copernic, 7000 Mons, Belgium.
 [6] Maarten Jansen, Hyeokho Choi, Sridhar Lavu, Richard Baraniuk, "Multiscale Image Processing Using Normal Triangulated Meshes", Dept. of Electrical and Computer Engineering Rice University Houston, TX 77005, USA.
 [7] S. Antani D. Crandall R. Kasturi, "Robust Extraction of Text in Video", Proceedings of the International Conference on Pattern Recognition (ICPR'00) 2000 IEEE.
 [8] Kobus Barnard and Nikhil V. Shirahatti, "A method for comparing content based image retrieval methods", Department of Computer Science, University of Arizona.
 [9] Rainer Lienhart and Frank Stuber, "Automatic text recognition in digital videos", University of Mannheim, Praktische Informatik IV, 68131 Mannheim, Germany.
 [10] Rainer Lienhart and Wolfgang Effelsberg, "Automatic Text Segmentation and Text Recognition for Video Indexing", ACM/Springer Multimedia Systems Magazine.
 [11] Jafar M. H. Ali About Ella Hassanien, "An Iris Recognition System to Enhance E- security Environment Based on Wavelet Theory", AMO - Advanced Modeling and Optimization, Volume 5, Number 2, 2003.
 [12] Jovanka Malobabiae, Noel O'Connor, Noel Murphy, Sean Marlow, "Automatic Detection And Extraction Of Artificial Text In Video", Adaptive Information Cluster, Centre for Digital Video Processing Dublin City University, Dublin, Ireland.
 [13] Chew Lim Tan, Ruini Cao, Qian Wang, Peiyi Shen, "Text Extraction from Historical Handwritten Documents by Edge Detection", School of Computing, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260.
 [14] Aurelio Velázquez and Serguei Levachkine, "Text/Graphics Separation and Recognition in Raster-scanned Color Cartographic Maps", Centre for Computing Research (CIC) - National Polytechnic Institute (IPN).

ACKNOWLEDGMENT

The work is carried out through the research facility at the Department of Electronics & Communication Engineering and Department of Electrical & Electronics Engineering, HITS (Holy Mary Institute of Technology & Science) College of Engineering, Bogaram(V), Keesara(M), R.R.District., HYDERABAD,A.P., Depts. Of , ECE/CSE from JNTUH/JNTUK, Dept., ECE/CSE from VJIT, Hyderabad. The Authors also would like to thank the authorities of JNTU for encouraging this technical work.

BIOGRAPHY



G.NAGENDRA is working as Asst.Prof. in Dept of ECE for Vidya Jyothi Institute of Technology college, HYDERABAD. He completed M.Tech (communications & signal processing) from S.K university, Anantapur. He completed B.Tech(ECE) from JNTU Hyderabad. His Areas of research includes DIGITAL SIGNAL PROCESSING, DIGITAL IMAGE PROCESSING and MOBILE COMPUTING.



Ms.Daggumati Rajani, working in VITS , Hyderabad, pursuing Ph.D from A.U, Vizag, M.Tech from JNTU Anantapur, B.Tech from NBKR, Vidyanagar. She has presented three papers from National Conferences. Research interest areas Digital Image Processing, Digital Signal Processing and Embedded Systems.



S. China Venkateswarlu, IEEE Member, B.Tech , M.Tech., MISTE, FIETE, IAENG., working as a PROFESSOR & HOD , from Holy Mary Institute Technology & Science , Bogaram(V), R.R. District, A.P, INDIA, Worked in various Engineering College and have 16 years of Teaching, and Industrial experience. Currently doing research in the area of Signal Processing / Speech Processing . He has presented papers under 10 National and 2 International Conference Papers are Presented through EDAs. He has published 15 research papers in National and International Journals. Research interest in Signal Processing, Speech Processing, Wireless Communications and Embedded Systems..



Mr. V.Sridhar, working as a **Assistant** Professor-ECE from VJIT-Hyderabad, M.Sc(IT) from Acharya Nagarjuna University, M.Tech from JNTUH, Hyderabad, He has having Life Member from ISTE, His interest area Digital Signal Processing Image Processing Mobile Comm and Wireless Communications .He has presented two technical papers from National Conferences.