ISSN:0975-9646

Morphological text localization using Wavelet and Neural network

M .Nagu, B.Raja Rao, B.R.B.Jaswanth
Department of Electronics and Communications Engineering, VKR & VNB Engineering College, Gudivada, Andhra Pradesh, S.India

Abstract:
Despite advances in the archiving of digital video, we are still unable to efficiently locate and retrieve the portions that interest us. Video indexing by segmentation has been a proposed solution and several research efforts are seen in the literature. Segmentation alone cannot solve the problem of content based access to video. Recognition of the text in video has been proposed as an additional feature. Various methods were proposed for isolation of the text data from the documented video image among wavelet transforms have been widely used as effective tool in the text segmentation.

This paper implements an efficient text algorithm for the extraction of text data from documented video clips. The implemented system carries out a performance analysis on various wavelets for the proper selection of wavelet transform with multilevel decomposition. Morphological operator is applied on the selected wavelet co- efficients for the text isolation and evaluates the contribution of decomposition levels and wavelet function to the segmentation results. This paper also implements neural networks for recognition of text characters from isolated text image and make tit editable.

Keywords: Morphological, wavelet, binarization, image

1. INTRODUCTION

Documents have been the traditional medium for printed documents. However, with the advancement of digital technology, it is seen that paper documents were 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 to 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. While electronic documents are best under storage of huge data base but very difficult for 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 academicia, business and industry. The 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.

Figure 1.1 Example of a documented video image clip

Text detection and recognition[2] in videos can help a lot in video content analysis and understanding, since text can

provide concise and direct description of the stories presented in the videos. In digital news videos, the superimposed captions usually present the involved person's name and the summary of the news event. Hence, the recognized text can become a part of index in a video retrieval system.The example of documented video image and extracted text data is shown in above fig1.1

2. RELATED WORK

Many efforts have been made for text extraction and recognition in video image sequence. Chung-Wei Liang and Po-Yueh Chen [1] in their paper DWT Based Text Localization presents an efficient and simple method to extract text regions from static images or video sequences. They implemented Haar Discrete Wavelet Transform (DWT) with morphological operator to detect edges of candidate text regions for isolation of text data from the documented video image.

A Video Text Detection And Recognition System presented by Jie Xi 1, Xian-Sheng Hua , Xiang-Rong Chen , Liu Wenyin , Hong-Jiang Zhang [2] proposed a new system for text information extraction from news videos. They developed a method for text detection and text tracking to locate text areas in the key-frames. Xian-Sheng Hua, Pei Yin , Hong-Jiang Zhang in their paper Efficient video text recognition using Multiple frame integration [3] presented efficient scheme to deal with multiple frames that contain the same text to get clear word from isolated frames.

C´eline Thillou and Bernard Gosselin proposed a thresholding method for degraded documents acquired from a low-resolution camera [4]. They use the technique based on wavelet denoising and global thresholding for nonuniform illumination. In their paper Segmentation-based binarization for color-degraded images [5] they described the stroke analysis and character segmentation for text segmentation. They proposed the binarization method to improve character segmentation and recognition.

S. Antani and D. Crandall in their paper Robust Extraction of Text in Video [7] describes an update to the prototype system for detection, localization and extraction of text from documented video images. Rainer Lienhart and Frank Stuber presented an algorithm for automatic character segmentation for motion pictures in their paper 'Automatic text recognition in digital videos' [9], which extract automatically and reliably the text in pre-title sequences, credit titles, and closing sequences with title and credits. The algorithm uses a typical characteristic of text in videos in order to enhance segmentation and recognition.

3. SYSTEM DESCRIPTION

The problem of text extraction [1] from video clip is divided into 4 main tasks namely wavelet decomposition, binarization, morphological operation and character recognition. The above description is shown with a system block diagram in fig 3.1

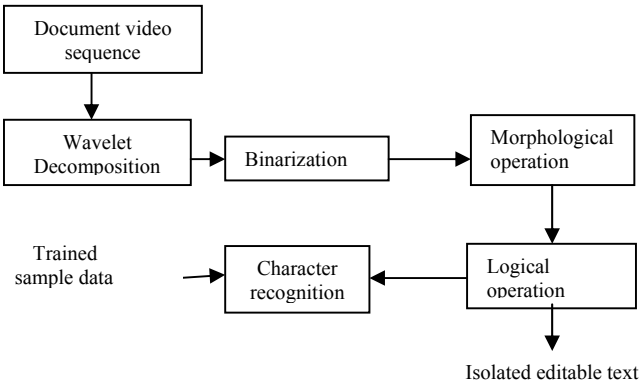


Figure 3.1 Block Diagram of the Implemented Design

3.1 *Wavelet decomposition:*

The documented video image sequences were considered for implementation consists of graphic and text. The wavelet transform is a very useful tool for signal analysis and image processing, especially in multi-resolution representation. It can decompose signal into different components in the frequency domain. One-dimensional discrete wavelet transform (1-D DWT) decomposes an input sequence into two components (the average component and the detail component) by calculations with a low-pass filter and a high-pass filter. The Two-dimensional discrete wavelet transform (2-D DWT) decomposes an input image into four sub-bands, one average component (LL) and three detail components (LH, HL, HH) as shown in Figure 3.2

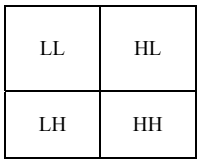


Fig.3.2 A two Dimensional DWT decomposition

In image processing, the multi-resolution of 2-D DWT [1] has been employed to detect edges of an original image. The traditional edge detection filters can provide the similar result as well. However, 2-D DWT can detect three kinds of edges at a time while traditional edge detection filters cannot. The traditional edge detection filters detect three kinds of edges by using four kinds of mask operators. Therefore, processing times of the traditional edge detection filters [3]is slower than 2-D DWT.The given fig.3.3 shows traditional edge detection using mask operation which consists of horizontal edge ,vertical edge and diagonal edge.

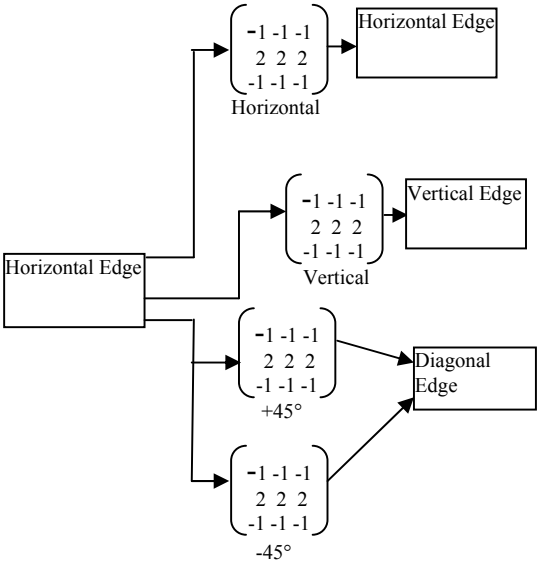


Figure 3.3 Traditional edge detection using mask operation

The proposed work implements three wavelet transforms namely Haar, Debuchies (orthogonal) and Spline Wavelet (Biorthogonal). The wavelet transform uses filter bank realization as shown in Figure 3.4 for the decomposition [3] of documented image into 3 details and 1 approximate coefficient. The 2-D DWT is achieved by two ordered 1-D DWT operations (row and column). Initially the row operation is carried out to obtain 1 D decomposition, then it is transformed by the column operation and the finally the resulted 2-D DWT is obtained. The 2-D DWT decomposes [1] a gray-level image into one average component sub-band and three detail component sub-bands as shown in Figure 3.4

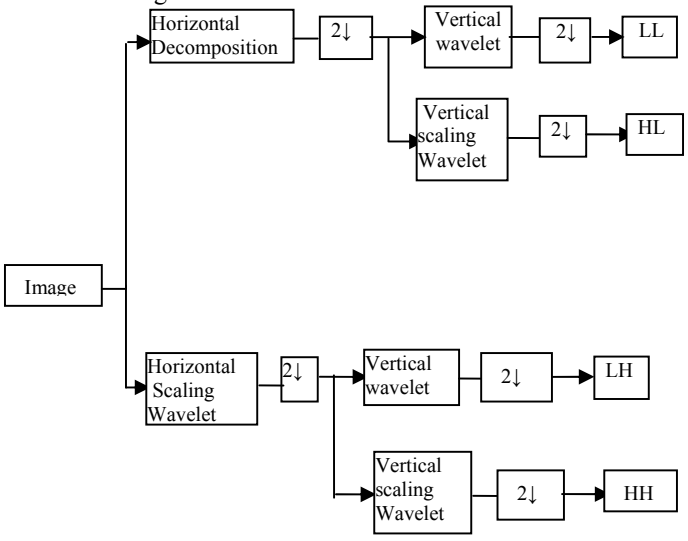


Fig.3.4 Filter Bank Implementation of Wavelet sub-band decomposition

The Fig 3.5 shows the original image used for the decomposition which is decomposed to 3 detail coefficients namely Horizontal, Vertical and Diagonal coefficients and 1 approximate coefficient as shown in Fig. 3.6. Which is a 1 level scaled image.

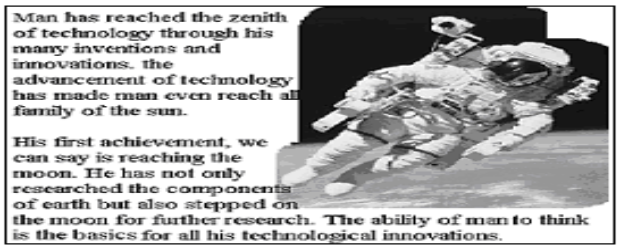


Figure3.5 Original Image

Figure **3.6** 1 level Scaled Image of the original image

3.2 *Binarization*

Binarization [5] is carried out using the thresholding. Thresholding [4] is a simple technique for image segmentation. It distinguishes the image regions as objects or the background. Although the detected edges are consisting of text edges and non-text edges in every detail component sub-band, they can distinguish due to the fact that the intensity of the text edges is higher than that of the non-text edges. Thus, an appropriate threshold can be selected to preliminarily remove the non-text edges in the detail component sub-bands.

A dynamic thresholding [4] value is calculated as the target threshold value *T*. The target threshold value is obtained by performing an equation on each pixel with its neighboring pixels. Two mask operators are used to obtain mask equation and then calculate the threshold value for each pixel in the 3 detail sub-bands. Basically, the dynamic thresholding method obtains different target threshold values for different sub-band images. Each detail component sub-band ***es*** is then compared with *T* to obtain a binary image (***e***).

The threshold *T* is determined by

$$T = \frac{\sum (es(i,j) \; X \; s(i,j))}{\sum s(i,j)} \qquad \text{-- ---(4.1)}$$

Where

$$s(i,j)=Max(\mid g1 * * es(i,j) \mid, \mid g2 * * es(i,j) \mid) \qquad \text{-----(4.2)}$$

and $g1= [-1\ 0\ 1]$, $g2=[-1\ 0\ 1]^t$ ---(4.3)

In Equation 4.2, "* *" denote two-dimensional linear convolution.

The given Fig 3.7 shows the example of a 4×5 detail component sub-band ***(es)***. The masked matrix element S(P8) is calculated as given in eqn 4.4

$$\begin{pmatrix} P1 & P2 & P3 & P4 & P5 \\ P6 & P7 & P8 & P9 & P10 \\ P11 & P12 & P13 & P14 & P15 \\ P16 & P17 & P18 & P19 & P20 \end{pmatrix}$$

Figure3.7 **4**×5 detail component sub-band *(es*

$$S(P8) = \max(|P9 - P7|,|P13-P3|) \quad \text{----------(4.4)}$$

And the similar operations can applied to each pixel, all S (i, j) elements can be determined for each detail component sub-band. Using Equation 4.1 threshold 'T' can then be computed, and the binary edge image (***e***) is then given by

$$\mathbf{e(i,j)}= \begin{cases} 255, \text{ if } es(i,j) > T \\ 0 \\ 1 \\ 2 \end{cases} \text{, otherwise} \quad \text{----------(4.5)}$$

The resulted binary image[5], as shown in Figure 3.8 mostly consisted of text edges with few non-text edges and binarized to 2 mean levels.

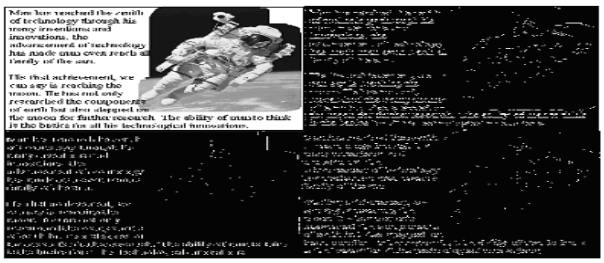


Figure 3.8Binary image of detail component sub-band

3.3 *Image dilation*

For the text region extraction [2], we use morphological operators and the logical operator to further remove the non-text regions. In text regions, vertical edges, Horizontal edges and diagonal edges are mingled together while they are distributed separately in non-text regions. Since determined text regions are composed of vertical edges, horizontal edges and diagonal edges, text regions can be to be the regions where those three kinds of edges are intermixed. Text edges are generally short and connected with each other in different orientation. Morphological dilation [8] and Erosion [9] operators are used to connect isolated candidate text edges in each detail component sub-band of the binary image. Figure 3.9 shows the Morphological operated scaled image.

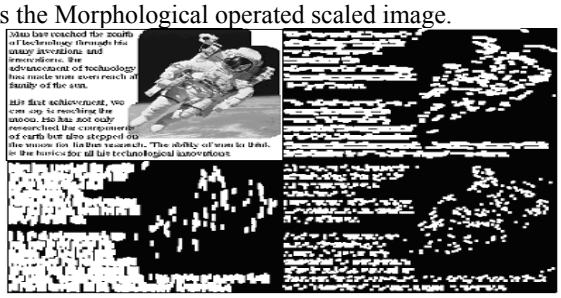


Figure3.9 Morphological operated image of three binary regions

The Morphological operators for the three detail sub-bands are designed differently so as to fit the text characteristics.

3.4 *Logical AND operation*

The logical AND [8] operation is then carried on three kinds (vertical, horizontal and diagonal) of edges after morphological operation to isolate the text data from the scaled image. Figure 3.10 demonstrates the application of the logical AND operator for the isolation of the Text Data in the operated Documented Image. The operator Performs Logical AND operation element by element for the three images and results the final ANDED Image. The Morphologically operated image have higher uniformity at text regions compared to the graphic region which results in elimination of the graphic region when ANDED.

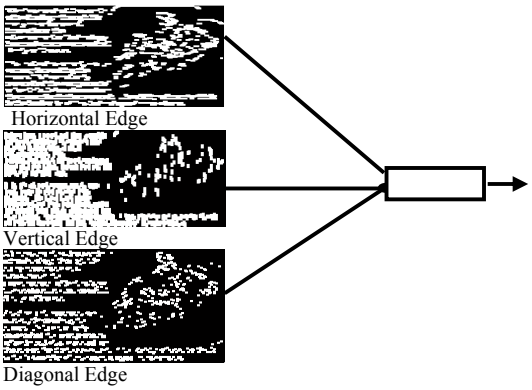


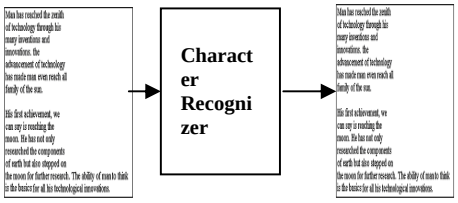
Figure 3.10 Text extraction by using the logical AND operator

Since three kinds of edge regions are intermixed in the text regions, overlapping appears a lot after the morphological operation due to the expansion of each single edge.,on the contrary, only one kind of edge region or two Horizontal Edge.

The kinds of edge regions exist separately in the non-text regions and hence there is no overlapping even after the morphologically operated. Therefore, the AND operator help in isolating the text regions as shown in Figure 3.10

3.5 *Character recognization*

The isolated characters are passed for recognition [3] to make it editable under character recognizer unit. The recognition of the character is carried out based on the trained value passed to the unit as shown in fig.3.11. The character recognition unit reads the isolated text data as test sample and performs the training operation to extract the feature for the test sample. To the extracted feature [10] of the test sample the minimum distance is calculated, the least minimum distance is assumed to be that character. The design estimates eleven features for the training of test sample and data base samples.



Extracted text data editable text data

Figure 3.11 character recognition operations.

The recognizer [3] extracts the texture features to distinguish between normal and abnormal characters. Four co-occurrence matrices are constructed in four different spatial orientations horizontal, right diagonal, vertical and left diagonal (0^0,45^0,90^0,135^0) as shown in Figure 3.12

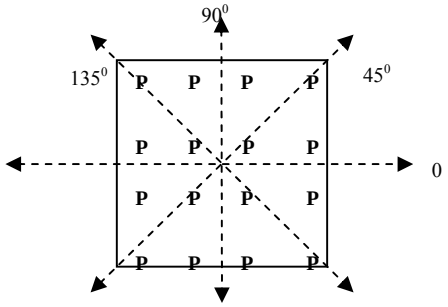


Figure 3.12

The features are extracted as presents in above section. The features are used for prediction of suitable characters from the data base depending on the minimum distance criterion.A set of eleven features are extracted in different orientations
for training of samples.

4. RESULTS AND REFERENCES

In this paper we have taken an avi video sequence as shown in fig 4.1 to produce results with four decomposition levels. We have also shown best result at fourth level decomposition with spine wavelet. The given video sequence consists of 5 frames and matrix size of 168x212.

Figure 4.1 an avi video sequence

4.1 *Text isolation using Haar wavelet & Debuchie wavelet*

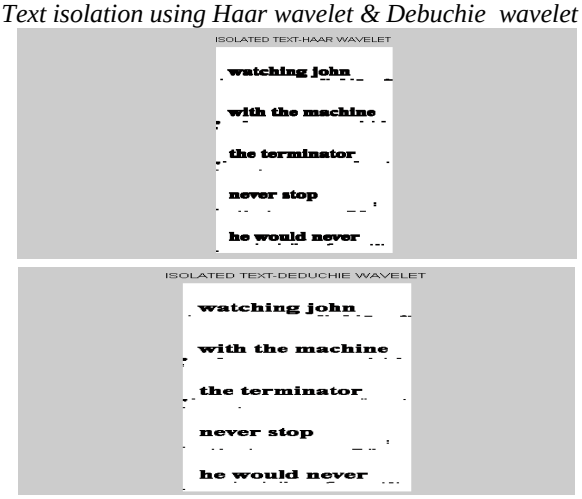


Figure4.2 Text isolation from video file using Haar wavelet and debuchies wavelet

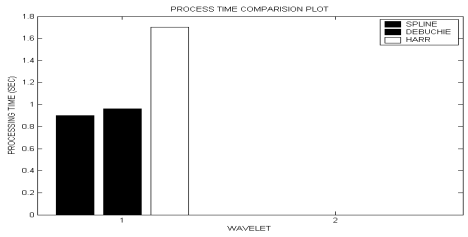


Figure4.3 Time comparison plot for WAVELETS

The above fig.4.2 shows the text, which is isolated from documented video data file using Haar wavelet transformation. and Debuchie wavelet transformation. The Fig.4.3 shows the time comparison plots for wavelets.

4.2 *Text isolation using Spline wavelet*

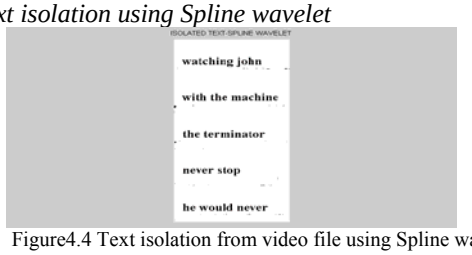


Figure4.4 Text isolation from video file using Spline wavelet

The fig.4.4 shows the text, which is isolated from documented video data file using Spline wavelet transformation.

4.3 *Error comparison plot for wavelets*

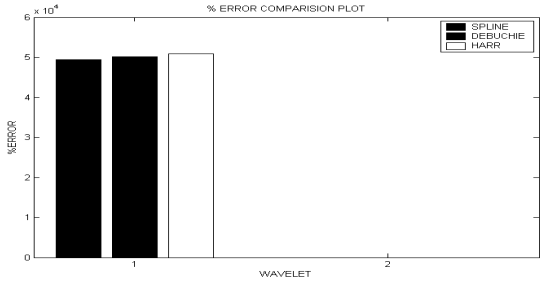


Figure 4.5 Error comparison plot for WAVELETS

The fig.4.5 shows the analysis of the % of error for multi level wavelets. It is observed that percentage of error is less in spline wavelet.

4.4 *Time comparison plot for wavelets*

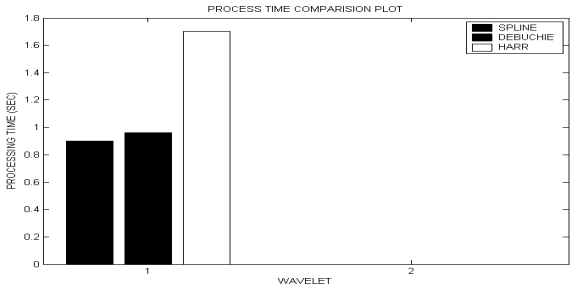


Figure 4.6 processing time comparison plot for WAVELETS

The work analyzes the efficient wavelet by comparing the time of processing taken by the Haar, Debuchie and Spline wavelets. The fig 4.6 shows the processing time comparison plot for wavelets(harr,spline and debuchie),and it is observed that the processing time is less in spine wavelet. The Documented video data is processed using various levels of scaling.(scale1,scale2,scale3 and scale4).

4.5 *Level 1 scaled image and isolated text from level 1 scaling*

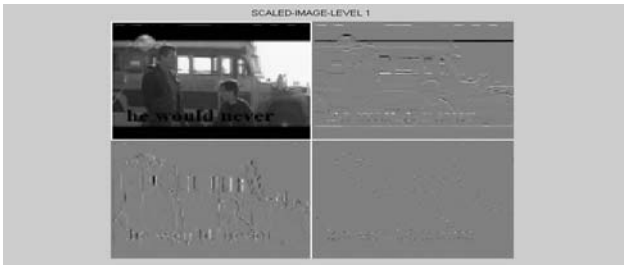


Figure 4.7 Level 1 scaled image

Figure 4.8 Isolated text from scale 1 image.

In the above Figures 4.7 and 4.8 shows the input documented video data file is viewed in 1 level scaled image and isolated text from that file.

4.6 Level 2 scaled image and isolated text from level 2 scaling

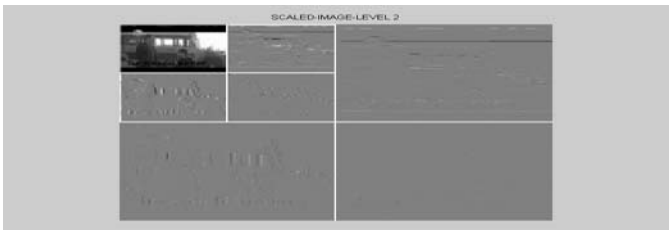


Figure 4.9 Level 2 scaled image

Figure 4.10 Isolated text from scale 2 image

In the above Figures 4.9 and 4.10 shows the input documented video data file is viewed in 2 level scaled image and isolated text from that file.

4.7 Level 3 scaled images and isolated text from level 3 scaling

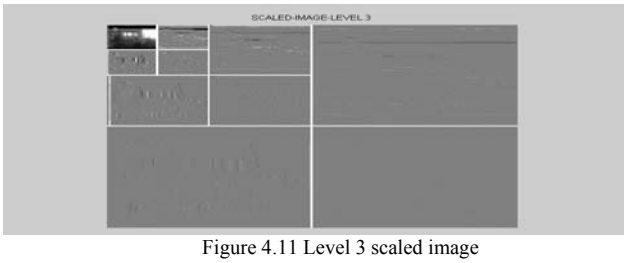


Figure 4.11 Level 3 scaled image

Figure 4.12 Isolated text from scale 3 image

In the above Figures 4.11 and 4.12 shows the input documented video data file is viewed in 3 level scaled image and isolated text from that file.

4.8 *Level 4 scaled images and isolated text from level 4 scaling*

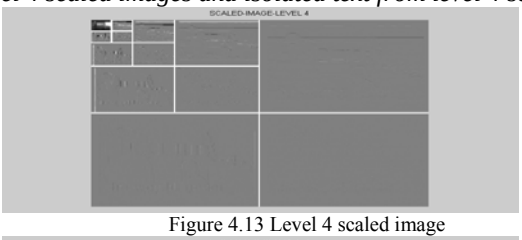


Figure 4.13 Level 4 scaled image

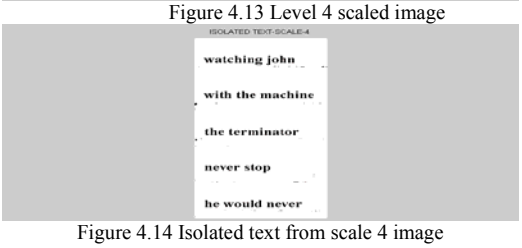


Figure 4.14 Isolated text from scale 4 image

In the above Figures 4.13 and 4.14 shows the input documented video data file is viewed in 4 level scaled image and isolated text from that file.

4.9 Error comparison plot for multi level scaling

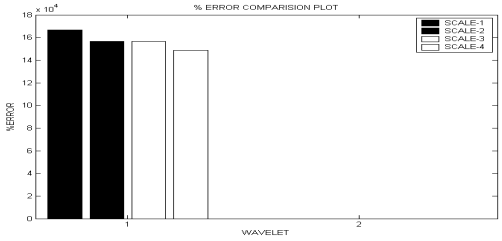


Figure 4.15 Error comparison plot for multi level scaling wavelets

The abovefig.4.15 analyzes the percentage of error for multi level scaling. It is observed that percentage of error is less in level 4 scaling.

4.10 Time comparison plot for multi level scaling

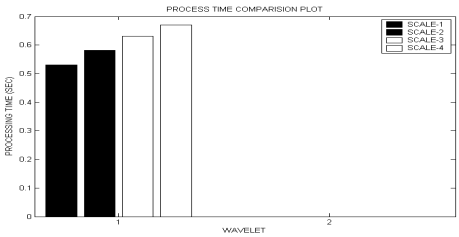


Figure 4.16 processing time comparison plot for multi level scaling wavelets

The work analyzes 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 plot for multilevel scaling is shown in fig 4.16.

4.11 Text isolation

Figure 4.17Text isolation using wavelet and multilevel scaling

The work analyzes 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. It is observed that the better result occurred at scale 4.The given fig.4.17 shows the extracted text data from the scale 4.

The given table I and table II shows the percentage error and computation time for multilevel scaling and multi wavelet analysis. The percentage error is less at scale 4 and process time is more .The percentage error is less with spline wavelet and comparison time is also less compared with other wavelets. Hence spline wave transformation is best for extraction of text.

Table I : %Error and computation time for Multi-Level Scaling Analysis

Scale	% error
Scale-1	16.2
Scale-2	15.8
Scale-3	15.79
Scale-4	14.2

Scale	*Process Time(Sec)
Scale-1	0.52
Scale-2	0.58
Scale-3	0.62
Scale-4	0.69

Table II: %Error and computation Time For Multi-Wavelet analysis

Wavelet	% error
Spline	4.96
Debuchie	5.05
Haar	5.1

Wavelet	*Process Time (Sec)
Spline	0.85
Debuchie	0.96
Haar	1.72

5. CONCLUSION

The project work realizes efficient text segmentation algorithm and character recognition for the isolated text data in a documented video image sequence. The text isolation system implements three wavelet transformations namely Harr, Debuchie and spline wavelet and analyzes the effect of these wavelet transformation on text isolation process for a given video sequence. The system also analyzes the effect of level decomposition on a documented video image sequence. The obtained results shows good isolation of text data form the image sequence for biorthogonal spline wavelet and shows better result at fourth level of decomposition. From the multi level and multi wavelet analysis the best suited wavelet transform and the best level of decomposition is obtained which is used under text isolation and its recognition. The spline wavelet transformation gives more accuracy to isolation of text data compared to haar and debuchie wavelet transform at higher level of decomposition. The implemented design also realizes the character recognition unit using supervised learning process. It is observed that the system recognizes the isolated text data form the video sequence to high accuracy.

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 Aboul 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 Malobabiæ, 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).