A Novel Method for Efficient Text Extraction from Real Time Images with Diversified Background using Haar Discrete Wavelet Transform and K-Means Clustering

Narasimha Murthy K N[1], Dr. Y S Kumaraswamy[2],
[1] Professor, Dept of Information Science and Engineering, V T U, City Engineering College, Bangalore, Karnataka, Pin-560062, India,

[2] Professor & HOD, Dept. of MAC (VTU), V T U, Dayananda Sagar College of Engineering, Bangalore, Karnataka, Pin-560078, India,

Abstract

The proposed system highlights a novel approach of extracting a text from image using two dimensional Haar Discrete Wavelet Transformation and K-Means Clustering. As the commercial usage of digital contents are on rise, the requirement of an efficient and error free indexing text along with text localization and extraction is of high importance. Majority of the previous research work on text extraction has focused on scene text, uniform background, and extensive use of wavelet domain and frequent usage of only grey-scale image as input. The extensive in-depth testing of such approach will lead no not-so-satisfactory results if the image type, non-uniform background, different text orientation, different languages are introduced. The proposed system has broader scale of consideration of input image with much complicated backgrounds along with consideration of sliding windows. For much accuracy, morphological operation is included to accurately distinguish the text and non-text area for better text localization and extraction. The experimental result was compared with all the prior significant work in text extraction where the results show a much robust, efficient, and much accurate text extraction technique.

Keyword: *Text Extraction, Haar, Discrete Wavelet Transform, K-Means Clustering, Morphological Operations*

1. Introduction

Text Extraction from images is a major task in computer vision. Applications of this task are various (automatic image indexing, visual impaired people assistance or optical character reading...). Many studies focus on text detection and localization in images. However, most of them are specific to a constrained context such as automatic localization of postal addresses on envelopes [1], license plate localization [2], text extraction in video sequences [3], automatic forms reading [4] and more generally ”documents” [5]. In spite of such extensive studies, it is still not easy to design a general-purpose TIE system [6]. This is because there are so many possible sources of variation when extracting text from a shaded or textured background, from low-contrast or complex images, or from images having variations in font size, style, color, orientation, and alignment. These variations make the problem of automatic TIE extremely difficult. Increasing popularity of digital cameras and camera phones enables acquisition of image and video materials containing scene text, but these devices also introduce new imaging conditions such as sensor noise, viewing angle, blur, variable illumination etc. Taking into account all these problems and scene text properties it is clear that its extraction and recognition is more difficult task in comparison with caption text and text in documents. Text information extraction consists of 5 steps [7]: detection, localization, tracking, extraction and enhancement, and recognition (OCR). In case of scene text particular focus is set on extraction. This step is done on previously located text area of image and its purpose is segmentation of characters from background that is separation of text pixels from background pixels. Text extraction strongly affects recognition results and thus it is important factor for good performance of the whole process. Text extraction methods are classified as threshold based and grouping-based. First category includes histogram-based thresholding [8], adaptive or local thresholding [9] and entropy-based methods. Second category encompasses clustering-based, region based and learning-based methods. Clustering techniques performed well on color text extraction [10]. Region-based approaches, including region-growing and split and merge algorithm, exploit spatial information to group character pixels more efficiently, but drawback is dependence on parameter values. Learning-based methods mostly refer to multi-layer perceptrons and self-organizing maps, but variation of scene text makes difficult to create representative training database.

The proposed work will introduce novel text extraction techniques with Discrete Wavelet Transform and k-Means Clustering. The system also introduces morphological operation like dilation and erosion for segregation of text and non-text regions for better accuracy. The rest of this paper is organized as follows. We discuss related work in Section II. The research methodology is discussed in Section-III. Proposed system is elaborated in Section IV. Implementation and Results is described in Section-V. Performance Analysis of the proposed system is discussed in Section-VI and finally conclusion and future work is described in Section-VII

2. Related Work

Syed Saqib Bukhari [11] presents a new algorithm for curled textline segmentation which is robust to above mentioned problems at the expense of high execution time. His approach is based on the state-of-the- art image segmentation technique: Active Contour Model (Snake) with the novel idea of several baby snakes and their convergence in a vertical direction only.

Samuel Dambreville [12] has combined the advantages of the unscented Kalman Filter and geometric active contours to propose a novel method for tracking deformable objects. Chen Yang Xu [13] have introduced a new external force model for active contours and deformable surfaces, which we called the gradient vector flow (GVF) field. The field is calculated as a diffusion of the gradient vectors of a graylevel or binary edgemap.

Wumo Pan e.t.al [14] has proposed a novel approach to detect texts from scene images captured by digital cameras. The system converts the text detection issue to a contour classification problem by means of the topographic maps, and performs shape classification by exploiting the over-complete and sparse structure in the shape data.

Fabrizio e.t. al [15] has presented a text localization technique which was considered to be efficient in the difficult context of the urban environment. The system uses a combination of an well-organized segmentation procedure based on morphological operator and a configuration of SVM classifiers with a variety of descriptors to estimate regions that are either text or non-text area. The system is competitive but generates many false positives Baba [16] has proposed a novel approach for text extraction by analyzing the textural evaluation in general scene images. The work has introduced a hypothesis that texts also have equivalent charecteristics that differentiates them from the natural background. The researcher has estimated spatial difference of texture to achieve the distribution of the degree of likelihood of text region.

Aghajari [17] propose an approach to automatically localize horizontally texts appearing in color and complex images. The text localization algorithm achieved a recall of 91.77% and a precision of 96%.

Hrvoje e.t. al [18] propose new method for scene text extraction in HSI color space using modified cylindrical distance as homogeneity criterion in region growing algorithm. The work has also introduced Solution for seed pixel selection based on horizontal projection.

Jayant e.t. al [19] present a novel method for extracting handwritten and printed text zones from noisy document images with mixed content. We use Triple-Adjacent-Segment (TAS) based features which encode local shape characteristics of text in a consistent manner. The experiment was tested with only similar types of text present in page. The system also lags different scripts testing.

Sumit [20] has presented a technique for using soft clustering data mining algorithm to increase the accuracy of biomedical text extraction. The development of the proposed algorithm is of practical significance; however it is challenging to design a unified approach of text extraction that retrieves the relevant text articles more efficiently. The proposed algorithm, using data mining algorithm, seems to extract the text with contextual completeness in overall, individual and collective forms, making it able to significantly enhance the text extraction process from biomedical literature.

3. Research Methodology

The issues of text extraction discussed in this proposed system from given image are multifold and can be segregated for various processing like binarization, implementing wavelet domain, morphological procedures, and finally localization and recognition of text.

The proposed system as shown in Figure 1 presents a research methodology where the text extraction from images with different scenario deploying discrete wavelet transform and k-means clustering. The prominent edges captured from the input binarized image are estimated using two dimensional discrete wavelet transform. Finally, when this stage is accomplished, morphological operations like erosion and dilation is implemented for the purpose of removing some non-text area which can be easily confused as text region. The morphological operations also associated various segregated candidate text regions in each information for sub-band of the binarized image. The fact in this stage for consideration is that binary information about the colors actually do not assist in text extraction procedure from the given image. The proposed system accepts input as colored RGB image for more real-time environment in development. The image is then processed in wavelet domain and then the text extraction process is implemented in later stage of processing. The proposed discrete wavelet transform system can be exhibited by following flow:

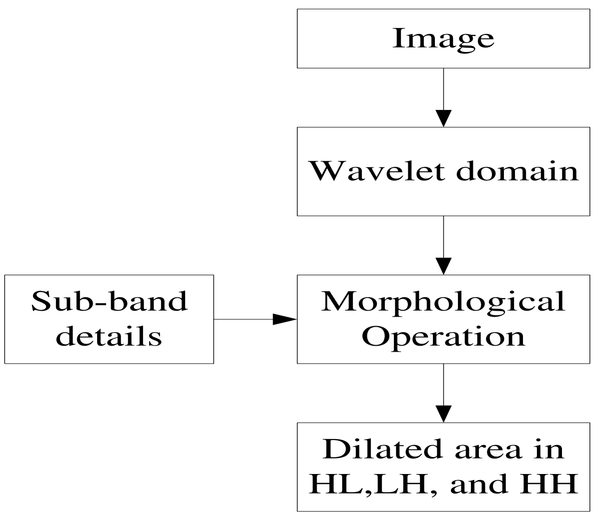


Fig 1. Proposed wavelet based text extraction protocol

A. Discrete Wavelet Transform

Digital image processing has witnessed a discrete wavelet transform as a prime tool in the area of multi-resolution analysis [21]. 1-D discrete wavelet transform decomposes an input image into mean constituent and detail constituent by estimation with the help of high-pass filter and low-pass filter [22]. Whereas 2D discrete wavelet transform will decompose an input image into 4 sub-bands (LL (*mean constituent*), LH, HL, and HH (*detailed constituent*)).

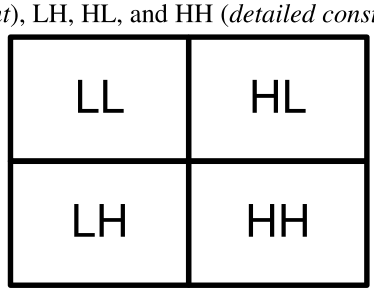


Fig 2. 2-D DWT decomposition output representation

The multi-resolution of the two dimensional wavelet domains can be deployed to explore the text regions of an input image. The conventional filters and detection mechanism for regions can also be expected to provide the equivalent output too. In comparison to one dimensional, 2D discrete wavelet transform can be the better option as it can identify maximum number of edges in one time which cannot be done by conventional algorithms. The conventional boundary detection filters can identify 3 types of boundaries using different types of masking operators as shown in Fig 3. This is also one of the significant reasons of why the conventional boundary detection filters are not faster in comparison to two dimensional discrete wavelet transform.

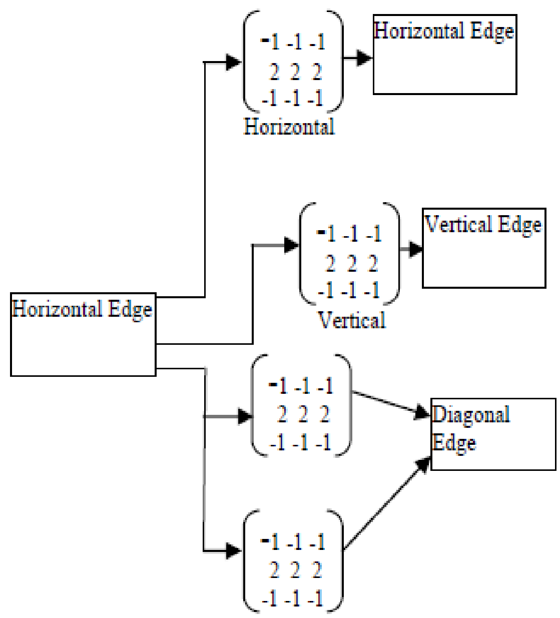


Fig 3. Conventional boundary detection by mask operator

(a) (b)

Fig 4. (a) Actual grey image (b) DWT coefficients

A grey scale image when achieved from the original input of RGB image is as shown in Fig 4(a). Fig 4(b) shows how the discrete wavelet transform converts the gray scale image into four sub-bands. The similar operation when performed by Haar [23] discrete wavelet transform makes the processing less complicated, faster with good accuracy, and efficient in comparison to other types of wavelet domain. The important features of the Haar wavelets are very contributing factors in the proposed methodology. The Haar DWT is genuine, symmetric, and orthogonal with simplest boundary situation along with support for random spatial grid distance. It also supports simple high-pass and low-pass filter coefficient [23].

$$\begin{bmatrix} A & B & C & D \\ E & F & G & H \\ I & J & K & L \\ M & N & O & P \end{bmatrix} \qquad \begin{bmatrix} (A+B) & (C+D) & (A-B) & (C-D) \\ (E+F) & (G+H) & (E-F) & (G-H) \\ (I+J) & (K+L) & (I-J) & (K-L) \\ (M+N) & (O+P) & (M-N) & (O-P) \end{bmatrix}$$

(a) (b)

$$\begin{bmatrix} (A+B)+(E+F) & (C+D)+(G+H) & (A-B)+(E-F) & (C-D)+(G-H) \\ (I+J)+(M+N) & (K+L)+(O+P) & (I-J)+(M-N) & (K-L)+(O-P) \\ (A+B)-(E+F) & (C+D)-(G+H) & (A-B)-(E-F) & (C-D)-(G-H) \\ (I+J)-(M+N) & (K+L)-(O+P) & (I-J)-(M-N) & (K-L)-(O-P) \end{bmatrix}$$

(c)

Fig 5. (a) The source image (b) Row operation in 2-D Haar DWT (c) Column operation in 2-D Haar DWT

A sample of 4x4 grey level images is shown in Fig 5(a). The addition and subtraction is applied on grey scale image for evaluating wavelet coefficient. The two dimensional discrete wavelet transform is accomplished by dual structured one dimensional discrete wavelet transform with both rows and columns. The row operation is conducted first in order to obtain the output as shown in Fig 5(b). Column operation is then used for transformation which finally gives the output of two dimensional Haar discrete wavelet transform as shown in Fig 5(c). A gray-scale image is converted to one mean constituent sub-band and three detail constituent sub-bands using two dimensional Haar DWT. Using Haar discrete wavelet transform on the image, diversified information about the text regions can be identified from the sub-bands details. For an example, LL subband identifies mean constituents, HL sub-bands identifies vertical boundaries, LH sub-bands identifies horizontal boundaries, and HH sub-bands identifies diagonal boundaries. The easy way to understand this is to observe the Fig 4 (a) which is basically a grey-scale image when subjected to Haar discrete wavelet transform gives the output as represented in Fig 5. The candidate text boundaries in the source image can seem from the detailed constituent's sub-bands (HL, LH, and HH).

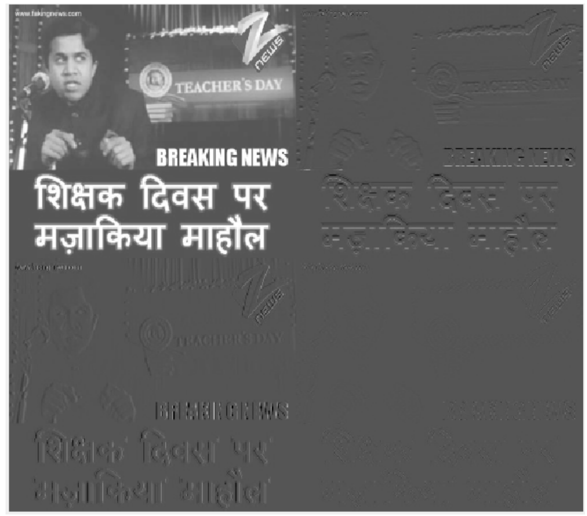


Fig 6. Implementing Haar discrete wavelet transform to source input image

B. K-Means Clustering

The k-means is basically a clustering algorithm which partition a data set into cluster according to some defined distance measure [24][25]. One of the significant tasks in machine learning is to comprehend images and extracting the valuable details. In this direction of analyzing data within the image, segmentation is the first phase to estimate quantity of the object present in an object. K-means clustering algorithm is an unsupervised clustering protocol [25] which categorizes the input data points into multiple types based on their inherent distance from each other. The protocol considers that the data features create a vector space and tries to locate normal clustering in them. The K-means function is given in (1).

[*mu*, *mask*]=kmeans (ima, k) (1)

where mu is the vector of class means, mask is the classification image mask, *ima* is the color image and *k* is the number of classes. The points are clustered around centroids in eq. (2) which are obtained by minimizing the objective [25].

Let m = max (ima)+1 , then

mu = {(1:k) * m} / (k+1) (2)

The maximum function shown above is the maximum value in the in *ima* matrix which represents the colored image in order to achieve the maximum value of the content colors where the color values are revealed as a unit value for all pixel. This stage is done to explicitly describe the maximum number of levels that can be used for estimating the histogram.

The working principle of the k-means clustering algorithm in the proposed system is as discussed below:

i. The histogram of intensities which should highlight estimates of pixels in that specific tone is estimated as shown below

$$n = \sum_{i=1}^{k} m_i \qquad (3)$$

where,

n = total estimates of observations
k = total estimates of tones.

The quantity of the pixels is estimated by the m_i which has equivalent value. The graph created with the help of this is only the alternative way to represents histogram.

ii. The centroid with k arbitrary intensities as in eq. (2) should be initialized.

iii. The following steps are iterated until the cluster labels of the image do not alters anymore.

iv. The points based on distance of their intensities from the centroid intensities are clustered.

v. The new centroid for each of the clusters is evaluated.

C. Morphological Operation

The morphological operations like dilation and erosions are used for better approach of refining text region extraction. The non-text regions are removed using morphological operations. Various types of boundaries like vertical, horizontal, diagonal etc are clubbed together when they are segregated separately in unwanted non-text regions. But, it is also known that the identified region of text consists of all these boundary and region information can be the area where such types of boundaries will be amalgamated. The boundaries with text are normally short and are associated with one other in diversified directions. The proposed system has deployed both dilation and erosion for associating separated candidate text boundaries in every detail constituent sub-band of the binary image.

Fig 7 Implementation of Morphological operations on three binary regions

Finally, the morphological operations like dilation and erosion is designed exclusively to fit use-defined input of text based image with various type of charecteristics.

4. Proposed System

The proposed work is designed to accept the input as an image where the final effective output is obtained as extracted text using k-means clustering algorithm and mathematical morphological operations. For contrast in the results, discrete wavelet transform is applied for decomposing the image to sub-bands at various scales with diversified resolution.

The text area is considered as special texture with unbalanced texture charecteristics. Various statistical features like mean, standard deviation, and energy is estimated when the image with text is subjected to discrete wavelet transformation algorithm. After the image is subjected to wavelet transform, classification based on region is applied for compacting the text area within the scope of image. A specific sliding window is designed which reads the high frequency sub bands by sliding steps. The application can be considered that the dimension of each sub-band is M×N after subjecting one-level wavelet transform, and we have,

$d_1 = \mathrm{mod}(M-W, l_1)$, $\quad d_2 = \mathrm{mod}(N-H, l_2)$

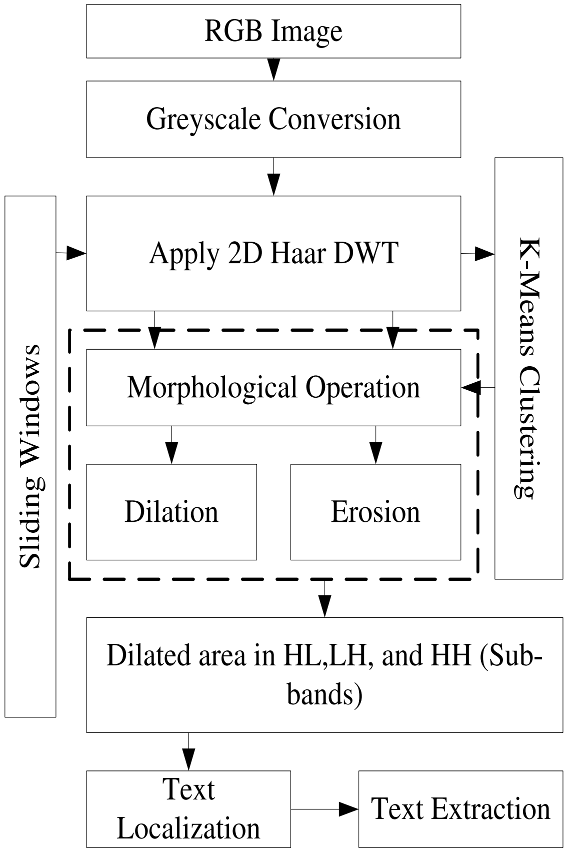


Fig 8. Overall Architecture of the proposed system

if d1 and d2 are not equal to zero, than it fails to superimpose all the area of every sub-band when sliding window reads the high frequency sub-bands by the step l_1 x l_2. The work also rejects all the contents which do not belong to the region.

The statistical charecteristics of every sub-band is estimated. The process achieves 12 features by evaluating the charecteristics of three high frequency subbands. Finally 12- dimension text feature vector is constructed.

The second phase of the design uses k-means clustering protocol where clustering is deployed by analyzing the texture characteristic vector. The clustering factors selected are primary point of text, normal background, and complex background. Care should be taken to update the point of cluster in every processing of k- iterations. The image is segregated into three categories for textual area, simple and complex background area. Binarization technique is applied to the image depending on the results of classification and then mathematical morphological operations are deployed to take out the text details from the image. The effective algorithm implemented in the proposed system is as follows:

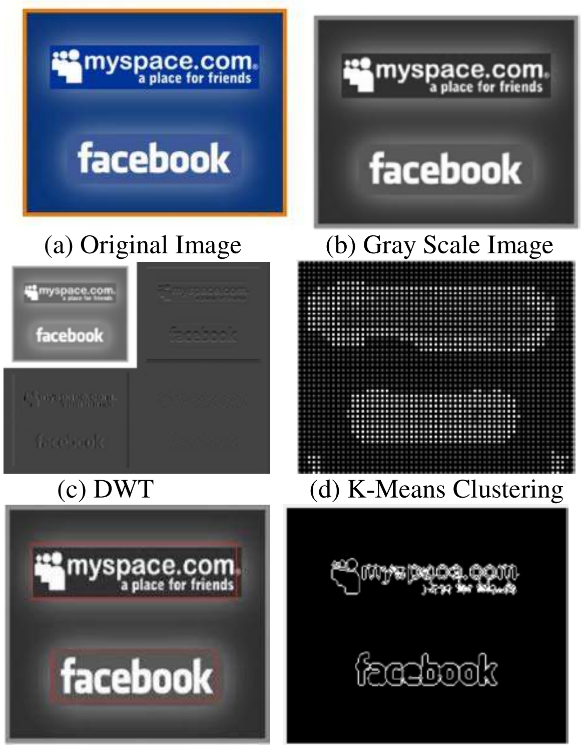
```
START
Input RGB image
If image is RGB
then covert to Gray scale
Create a function for performing DWT
Use Haar 2D DWT
Perform DWT
Initialize the coefficients, sub-bands
Create a function for sliding window
[W H] =size (window1)
mu = mean (mean (window1))
window2 = (window1-mu)
stanDev= sqrt (sum (sum (window2.^2))/(W*H))
E = sum (sum (window1.^2))
Estimate Size of subband
Create a function for K-Means Clustering
Calculate column number and row number
For zero padding
Apply zero Padding
Extract the features of sliding window
Rebuild the cluster id
Apply Mask Operation
Morphological operations on binary images
Detect boundary using Sobel
Morphologically open binary image
      (remove small objects)
STOP
```

One of the prime issues of implementing clustering algorithm is an inevitable computing error for which reason once the text area is extracted, the system cannot facilitate wholesome error free information about the complete text area. Therefore, the design implements morphological operations like erosion and dilation in order to measure and localize the all text sub-areas. Another issue is the non-text pixels which are also eliminated using erosion and dilation. The appropriate position of the text region is localized in the original image by merging the text pixel locus that is not extracted around the text region boundary. Finally the actual text information is extracted from the processed binarised image.

5. Implementation and results

The framework project work is designed in Matlab in 32 bit system 1.8 GHz with dual core processor where total of 150 different types of images are considered for the experiment. The basic graphics video display card of DIAMOND AMD ATI Radeon is used for experimenting on both OS of Windows Vista and Windows 7. The implementation also considers images with single text, multiple text, text with different sizes of fonts, text with complex and simple background, text with different languages.

The input image binarised to grayscale which is then subjected to discrete wavelet transform. The system then subjects the processed image into k-means clustering protocol. Morphological operation like erosion and dilation is deployed in order to remove all the unwanted non-text region which can be confused with the text regions sometimes. Finally text localization and extraction takes place as shown in the results below:



(a) Original Image (b) Gray Scale Image
(c) DWT (d) K-Means Clustering
(e) Erosion/ Dilation (f) Text Localization

Fig 9. Results from Text Extraction Process

The above results in Fig 9. shows the output of the application obtained when an image of simple background and different multiple text with different font size is used. The localization process along with text extraction is found to be satisfactory.
Our preliminary experiment although was not so satisfactory when the attempt was conducted on the scanned image for text extraction. The accuracy rate was only 75%. The experiment is also conducted in image with text in Hindi language unlike the previous experiment. To identify the robustness and compatibility of the designed application, the experiment was conducted with two set of image e.g.:

- Image with Hindi text with simple background and with same font size.
- Image with both Hindi and English text with different font size and style and orientation.

The second set of the experiment is conducted to scrutiny the efficiency of the protocol towards text extraction for non-English text. Here we chose Hindi language for testing as it is one of the most frequently used language in any type of document related to Indian Government. Fig 10. shows the reliability of the application for extracting the Hindi text. The error percentage is zero in this case showing system to be robust in Hindi Language too along with English. But a fact has to consider that this experiment is conducted with condition of simple background and not all the text will have simple background.

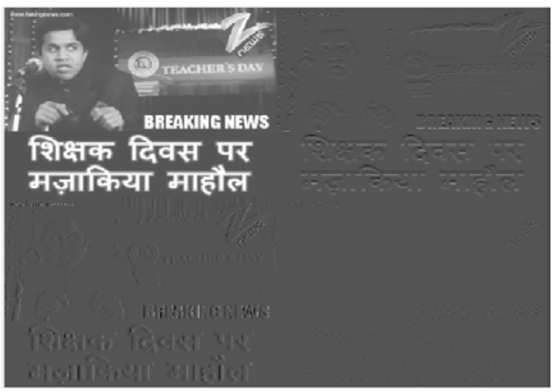
Fig 10. Output of Text Extraction for Hindi Font with simple background

The third set of the performance analysis is conducted considering complex background. Complex background can be defined as an image with high variation of RGB along with illumination factor in its background whereas in simple background it is uniform. Therefore, it was a bit of challenging task to have proper consideration of image with multiple text of different font as well as with complex background. So for this set of experiment, we have selected an image captured from the running live video streaming from using TV tuning card. A good graphics adapter will be required for proper restoration of the captured image.
The image for this set of experiment is considered as an image with:

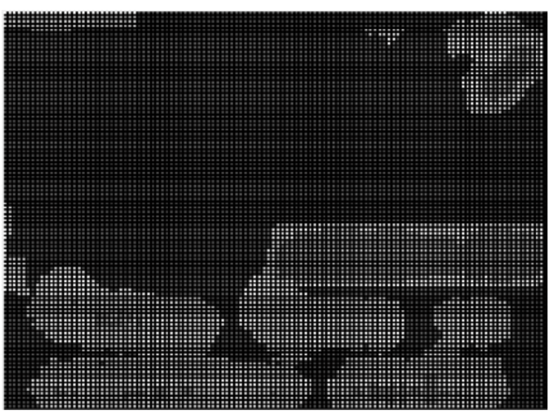
- Multiple Text
- Multiple Text with different font size

- Multiple text with different language and its orientation.

(a) Binarized image



(b) Applying DWT to the Binarized image



(c) After implementing k-Means Clustering Protocol

(d) Morphological Operation implements to remove non-text regions

(e) Text Region Localization

(f) Text Extraction
Fig 11. Text Extraction in complex background

It can be noticed here that almost all the text (both in Hindi and English) is 100% accurately extracted in all configurations of the sliding window. Interesting fact is the English text on the top right side of the channel logo which has different orientation or inclination in comparison to other text is also extract with 95.2 % accuracy. Therefore, the system design for the proposed text extraction can be eventually considered as robust, reliable, and efficient in text extraction in multiple scenarios.

6. Performance Analysis

For evaluation purpose, the proposed system deploys actual rectified size and error detected size to analyze the simulation results accomplished.

Actual Rectified Size = $(A_R / A_T) \times 100\%$
Error Detected Size = $(N_T / (A_R + N_T)) \times 100\%$

Where
A_R = Actual Extracted Text Region Size
N_T = Non-Text Region Size
A_T = Actual Text Region Size

The analysis results are optimum when actual rectified size is greater than error detected size. The experiment when evaluated with total of 150 images gives the following results as depicted in Table 1.

Table 1: Analysis Results

H	W	Error Detected Size	Actual Rectified Size
16	32	13.6	94.5
16	16	8.5	91.8
8	16	5.1	88.6
8	8	4.8	86.3

The above results interpret that using the proposed algorithm assist the applications to achieve reduction in error detection size and higher actual rectified size. The process successfully extracts the text from various sets of experimental images with the sliding window size of smaller dimension.

Comparison with Techniques

Majority of the research work conducted in past has used datasets of images. But the proposed work is totally focused on real time images being captured from digital camera, or mobile phone, or from any image capturing devices for better study for realistic result. The current research work has been compared with certain conventional algorithm for cross-checking its efficiency. All the analysis process employs the exploration of accuracy of text and non-text area for the given colored image input. The method for checking performance has deployed the complex methodology of discovering all the prominent boundaries and contours at different orientations considering images with multiple text with different font size, style, and language (English, Hindi) for the proposed process of text extraction. It has been noticed that the complexity of the applied protocol increases in order to recognise the boundaries at multiple different directions.

The prominent morphological operator like dilation is employed for sequencing the clusters of the segregated text to a significant complete word or sentences. The proposed system also explores successfully the quantity of the constituents and estimates the degree of inconsistency for each constituent variance. The consideration of detection of text is symbol in case the value of the inconsistency of each constituent is greater. The proposed algorithm though it was found to be very sensitive to skew and direction of placement of text, but the result accomplished in majority of test on 150 images were found to be successful. One of the prime intentions behind the proposed text extraction system is to diminish the probability of detection of non-text elements from the test image. The effectiveness of the morphological operations was also tested by analyzing their respective output image.

The proposed system has also being compared and experimented with all the major significant previous research work like:

- Morphological approach considering scene text by Hasan and Karam (2000) [26]
- Wavelet based feature extraction and neural network for texture analysis considering slanted scene text, localization, and tracking by Li et. al (2000) [27]
- Text detection and localization using DCT coefficient and macroblock type information by Lim et. al (2000) [28]
- Gabor filter like multi-layer perceptron for texture analysis by Jung e.t. al (2001) [29]
- Text detection using sparse representation by Pan et. al (2009) [14]
- Text Localization algorithm in color image via New projection profile by Aghajari (2010) [17]
- Text extraction using data mining algorithm by Sumit (2011) [20]

All the above research work done has extensively used the thresholding concepts along with morphological operations. Whereas the proposed system has contrast implementation of the above mentioned work along with novel introduction of k-means clustering and Haar discrete wavelet transform of two dimensional. All the 150 images has been tested with the above mentioned research work and compared with the proposed work to observe that majority of the experiments with the previous approach when used, it gives better results only in case of consistent background. One more prime observation is that when the images with multiple text style, size, orientation and especially languages are used, all the previous approaches yields false results of text extraction. For example, when the above mentioned research works is implemented on the same test image, the results obtained are as followings:

Fig 12(a)

Fig 12(b)

The issue observed above is for the image with uniform background (Fig 12(a)) the text localization is somewhat valid, but it fails to locate the text when background is making inconsistent like in Fig 12(b). There is much such type of errors in the results obtained when previous approaches are compared with the existing one. The proposed protocol for text extraction is therefore considered to be robust and efficient.

7. Conclusion

The proposed system has introduced a novel process of text extraction considering multiple cases of image with its textual contents. The system has been implemented using 2D Haar DWT along with k-means clustering algorithm. It also deploys methodology of sliding window for reading sub-bands of high frequency. Morphological operations like dilation and erosion has been introduced finally to refine the text and non-text region appropriately. For more realistic and robust results, the proposed system has been experimented with images with single / multiple text, multiple text of different sizes / style / languages, images with uniform and non-uniform background. The system is also evaluated with major research results in the past for conventional text extraction approach and is found to be potential for more accurately extracting text information. The future work will be to extending the similar concept of extracting text from video with higher accuracy.

Reference

[1] Palumbo, P. W., Srihari, S. N., Soh, J., Sridhar, R. and Dem-janenko, V., 1992. Postal address blocks location in real time. Computer 25(7), pp. 34–42.

[2] Arth, C., Limberger, F. and Bischof, H., 2007. Real-time license plate recognition on an embedded DSP-platform. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR '07) pp. 1–8.

[3] Wolf, C., Michel Jolion, J. and Chassaing, F., 2002. Text localization, enhancement and binarization in multimedia documents. In: In Proceedings of the International Conference on Pattern Recognition (ICPR) 2002, pp. 1037–1040.

[4] Kavallieratou, E., Balcan, D., Popa, M. and Fakotakis, N., 2001. Handwritten text localization in skewed documents. In: International Conference on Image Processing, pp. I: 1102–1105.

[5] Wahl, F., Wong, K. and Casey, R., 1982. Block segmentation and text extraction in mixed text/image documents. Computer Graphics and Image Processing 20(4), pp. 375–390.

[6] Keechul Junga, Kwang in Kimb, Anil K. Jainc, Text information extraction in images andvid eo: a survey, Pattern Recognition 37 (2004) 977 – 997

[7] H.J. Zhang, Y. Gong, S.W. Smoliar, S.Y. Tan, Automatic parsing of news video, Proceedings of IEEE Conference on Multimedia Computing and Systems, Boston, 1994, pp. 45–54.

[8] A.W.M. Smeulders, S. Santini, A. Gupta, R. Jain, Content-basedimage retrieval at the endof the early years, IEEE Trans. Pattern Anal. Mach. Intell. 22 (12) (2000) 1349–1380.

[9] M.A. Smith, T. Kanade, Video skimming for quick browsing basedon audio andimage characterization, Technical Report CMU-CS-95-186, Carnegie Mellon University, July 1995.

[10] M.H. Yang, D.J. Kriegman, N. Ahuja, Detecting faces in images: a survey, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 34–58.

[11] Syed Saqib Bukhari, Coupled Snakelet Model for Curled Textline Segmentation of Camera-Captured Document Images, Proceedings of the 2009 10th International Conference on Document Analysis and Recognition, 2009

[12] Samuel Dambreville, Tracking deformable objects with unscented Kalman filtering and geometric active contours, American Control Conference, IEEE, 2006

[13] Xu. C, Prince. J, "Snakes, shapes, and gradient vector flow" , IEEE Transaction on Images Processing, Vo!. 7, 1998, pp. 359-369

[14] Wumo Pan, T.D. Bui, C.Y. Suen, Text detection from natural scene images using topographic maps and sparse representations, IEEE, 2009

[15] J. Fabrizio, M. Cord, B. Marcotegui. Text Extraction from Street Level Images, CMRT09 - CityModels, Roads and Traffic 2009. Paris, France.

[16] Baba, Y.[Yoichiro], Hirose, A.[Akira], Spectral Fluctuation Method: A Texture-Based Method to

Extract Text Regions in General Scene Images, IEICE(E92-D), No. 9, September 2009, pp. 1702-1715
[17] G. Aghajari, J. Shanbehzadeh, and A. Sarrafzadeh, A Text Localization Algorithm in Color Image via New Projection Profile, Proceedings of The International Multi conference of Engineers and Computer Scientist, Vol-2, 2010
[18] Hrvoje Dujmić, Matko Šarić, Joško Radić, Scene text extraction using modified cylindrical distance, Proceeding NNECFSIC'12 Proceedings of the 12th WSEAS international conference on Neural networks, fuzzy systems, evolutionary computing & automation, World Scientific and Engineering Academy and Society (WSEAS), 2011
[19] Jayant Kumar; Rohit Prasad; Huiagu Cao; Wael Abd-Almageed; David Doermann; Premkumar Natarajan, Shape codebook based handwritten and machine printed text zone extraction, Proceedings Vol. 7874, Document Recognition and Retrieval XVIII, 2011
[20] Sumit Vashishta, Yogendra Kumar Jain, Efficient Retrieval of Text for Biomedical Domain using Data Mining Algorithm, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 4, 2011
[21] R. Pradip Kumar, P, Nagabhushan, Multiresolution Knowledge Mining using wavelet transform, Engineering Letter, 2007
[22] Abhayaratne, G.C.K., Discrete wavelet transforms that have an adaptive low pass filter, Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on Issue Date: 1-4 July, Vol 2, 2003
[23] Patrick J. Van Fleet, Discrete Haar Wavelet Transforms, PREP - Wavelet Workshop, 2006

[24] S K Gupta, K Sambasiva Rao and Vasudha Bhatnagar, K-means Clustering Algorithm for Categorical Attributes, DataWarehousing and Knowledge Discovery Lecture Notes in Computer Science, 1999
[25] Hassana Grema Kaganami, Zou Beiji, M Sami Soliman, Optimal Color Image Enhancement Using Wavelet and K-means Clustering, International Journal of Digital Content Technology and its Applications. Volume 5, Number 1, January 2011
[26] Y.M.Y. Hasan, L.J. Karam, Morphological text extraction from images, IEEE Trans. Image Process. 9 (11) (2000) 1978–1983.
[27] H. Li, D. Doerman, O. Kia, Automatic text detection and tracking in digital video, IEEE Trans. Image Process. 9 (1) (2000) 147–156.
[28] Y.K. Lim, S.H. Choi, S.W. Lee, Text extraction in MPEG compressed video for content-based indexing, Proceedings of International Conference on Pattern Recognition, 2000, pp. 409–412.
[29] K. Jung, Neural network-basedtext location in color images, Pattern Recognition Lett. 22 (14) (2001) 1503–1515.

Mr. K N Narasimha Murthy is an experienced teacher for more than two decades and presently working as a Professor and HOD in department of Information Science and Engineering at City Engineering college Bangalore, Affiliated to Visvesvaraya Technological University, India. He got BE degree in 1989, M.Tech degree in 1993, MCA degree in 2003 and M.Tech degree in 2006. His areas of interest are Algorithms, Image processing and Computer Vision and also a member of MIE, MISTE, IACSIT.

Dr.Y S Kumara swamy working presently as Sr. professor and HOD, Department of MCA (VTU) at Dayandasagar college of Engineering Bangalore, India. He has published more than 150 research papers and guided 38 PhD students in the area of Computer Science/ Mathematics and also a selection committee member for ISRO/UGC/DSI profiles