International Journal of Science Engineering and Advance Technology

Character segmentation from multi-oriented video words using Discrete Wavelet Transform and K-Means Clustering

B.Jagadeesh babu, M. Tech scholar, D.Kishore, Associate Professor

Abstract

The paper presents a two stage methods for the image segmentation from multi-oriented video words by using of DWT process 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 data has been taken from a d.w.t for compression of the data from the output of gray scale . 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, Discrete Wavelet Transform, K-MeansClustering, Morphological Operations, sobel technology*

1. Introduction

Text present in videos plays an important role in video indexing and retrieval and video text has been classified into two groups namely *'Scene text'* (e.g. text on vehicles, commodities, buildings, sign boards on roads, etc.) and *'Graphic text'* or *'Caption text'* (news video, sports video, etc). Hence both types of text can be extremely useful in the effective indexing and retrieval of the videos. The major challenges in text information extraction from video are low resolution, complex/non uniform backgrounds and blur, to mention a few.

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...). 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:[1] 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[2], adaptive or local thresholding[3] and entropy-based methods. Second category encompasses clustering-based, region based and learning-based methods. Clustering techniques performed well on color text extraction.[4] 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.

2. 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.

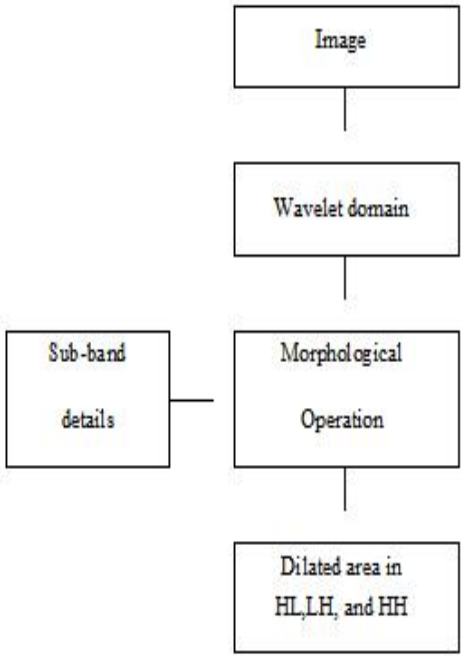

Fig 1. Proposed wavelet based text extraction protocol

A. Discrete Wavelet Transform

Digital image processing has witnessed a discrete wavelet transform as a prime tool in the area of multi-resolution analysis. 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 Whereas 2D discrete wavelet transform will decompose an input image into 4 sub-bands (LL (*mean constituent*), LH, HL, and HH (*detailed constituent*)).

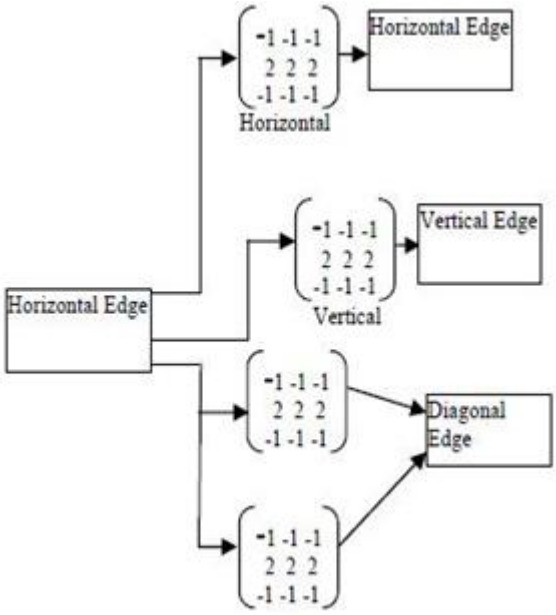

fig 3.Conventional boundary detection by mask operator

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.

(a) (b)

Fig 4. Actual grey scale Image DWT Coefficients

c) A grey scale image when achieved from the originalinput ofRGB image is as shown in Fig 4(a). Fig 4(b) shows how the discretavelet transform converts the gray scale image into four sub-bands.The similaroperation when performed by discrete wavelet transform makes the processing less complicated,faster withgood accuracy, and efficient in comparison to other types of wavelet accuracy, and efficient in comparison to other types of wavelet domain.The important features of the wavelets are very contributing factors in the proposed methodology.The 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.

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

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

A gray-scale image is converted to one mean constituent sub-band and three detail constituent sub-bands using two dimensional DWT. Using 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-image when subjected to 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 H

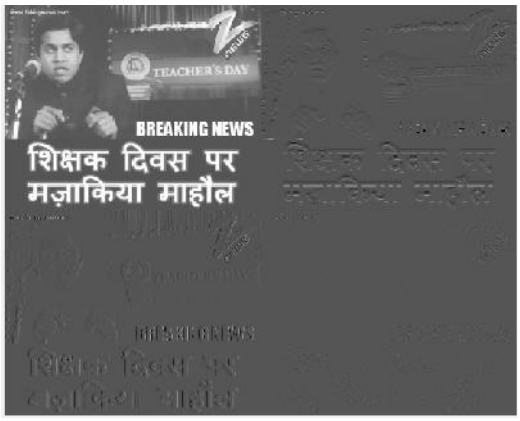

Fig 6. Implementing discrete wavelet transform to source input im

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

i=1

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,

d1= mod (M-W, l1), d2= mod(N-H, l2)

RGB Image

Grey scale Conversion

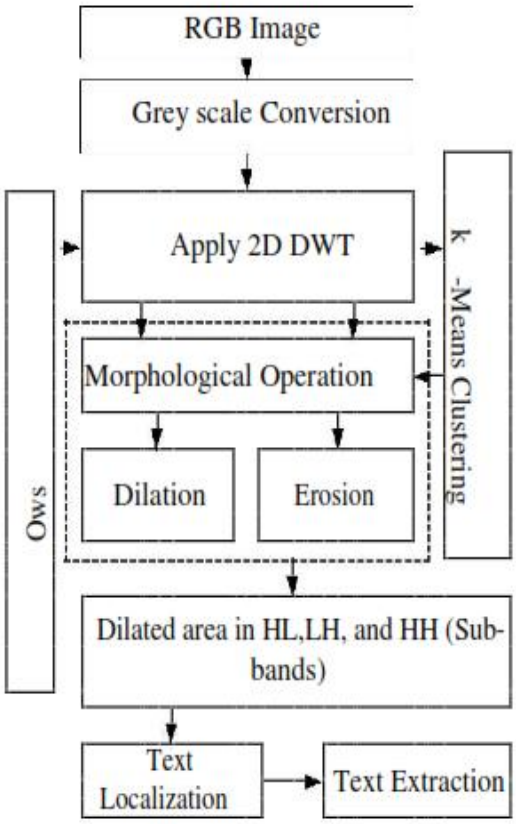

Fig 8. Overall Architecture of the proposed system

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
1 Input RGB image
2 If image is RGB
3 then covert to Gray scale
4 Create a function for performing DWT
5 Use 2D DWT
6 Perform DWT
7 Initialize the coefficients, sub-bands
8 Create a function for sliding window
9 [W H] =size (window1)
10 mu = mean (mean (window1))
11 window2 = (window1-mu)
12 stanDev= sqrt (sum (sum (window2.^2))/(W*H))
13 E = sum (sum (window1.^2))
14 Estimate Size of subband
15 Create a function for K-Means Clustering
16 Calculate column number and row number
17 For zero padding
18 Apply zero Padding
19 Extract the features of sliding window
20 Rebuild the cluster id
21 Apply Mask Operation
22 Morphological operations on binary images
23 Detect boundary using Sobel
24 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 binaries' 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:
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 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 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:

Conclusion

This method for multi-oriented video character segmentation has been proposed in this paper. The proposed method is a two-stage approach, where in the first stage the isolated (non-touching) characters are segmented, and in the second stage the touching characters are segmented

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 DWT along with k-means clustering algorithm. It also deploys methodology of sliding window for reading sub-bands of high frequency.

REFERENCE

[1] 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.

[2] 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.

[3] 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.

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

[5] 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

[6] 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

[7] 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

[8] 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

[9] 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

[10] 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

[11] 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

[12] 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. Document Recognition and Retrieval XVIII, 2011

[13] 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