

International Journal of Engineering Sciences & Research Technology

(A Peer Reviewed Online Journal)
Impact Factor: 5.164



Chief Editor
Dr. J.B. Helonde

Executive Editor
Mr. Somil Mayur Shah

ABSTRACT

The Intelligent content extraction framework consequently distinguishes and removes the content present in various sorts of pictures. The development of advanced world Detection and Extraction of content locales in a picture are notable issues in the zone of picture handling. The development of computerized world and the use of mixed media produced another period with an exemplary issue of example acknowledgment. Along these lines Automatic content extraction from pictures and recordings serves a significant job in ordering and effective recovery of media. The current systems, for example, area based, surface based strategies for the content extraction are not ready to smaller with every one of the uses of content extraction. The proposed Intelligent Text Extraction framework consequently recognizes and removes the content present in various sorts of pictures. The framework comprises of various stages which incorporate the restriction, division, extraction and acknowledgment of content from the pictures. In the proposed framework we utilize the discrete wavelet change for the restriction of content. The morphological activities are utilized which upgrades the recognizable proof of right content parts. The content part is divided and is perceived utilizing a productive framework. The upside of the framework is that the removed content is appeared in the .txt record. The proposed framework additionally permits the alteration of the perceived content from the picture. This strategy shows better effectiveness, exactness and review contrasted with the current strategies. This demonstrates the likelihood of utilizing this strategy in progressively new and propelled applications.

KEYWORDS: DWT, image processing, morphological operations, Otsu, segmentation, text extraction, text recognition

1. INTRODUCTION

The fast progression in the innovation and interactive media has digitalized the world. The accessibility of cameras and different frameworks contributes huge number of pictures to the world. Going from cameras implanted in cell phones to proficient ones, Surveillance cameras to communicate recordings, consistently pictures to satellite pictures, all these adds to increment in multi-media information. The vast majority of the pictures may contain message as a component of it, which gives some data about that picture. Subsequently distinguishing proof of these writings has importance in numerous applications. This demonstrates the significance of the content extraction framework in part of uses. It was expressed that as of late there was a radical in-wrinkle in sight and sound libraries and the measure of information is developing exponentially with time. It is likewise realized that there are number of TV channels that communicate regular and the far reaching of moderate computerized cameras and cheap memory gadgets, the multi-media information is expanding each second. The content present inside a picture empowers applications, for example, catchphrase based picture search, content based picture ordering and programmed video logging. Extraction of these writings from pictures is an exceptionally troublesome assignment because of varieties in character textual styles, styles, sizes and content bearings, and nearness of complex foundations and variable light conditions. By and large, the pictures can be classified into three dependent on its sort: report pictures, scene pictures and conceived computerized pictures. Archive pictures are the picture of the record which incorporates pdf, notes and so forth. The content present in record picture is the archive content.

Fig.1.(a) shows the document image.

Intelligent text extraction system

The Intelligent text extraction system automatically identifies and extracts the text present in different types of images. The growth of digital world Detection and extraction of text regions in an image are well known problems in the area of image processing. The growth of digital world and the usage of multimedia generated a new era with a classic problem of pattern recognition. Thus Automatic text extraction from images and videos serves an important role in indexing and efficient retrieval of multimedia.

(a)

Fig.1. (a) Document image



(b)

Fig.1. (b) Scene text image

Size: The text can have variable size from small to large.

(a) **Alignment:** The characters in caption text may appear in clusters and sometimes they can also appear as non-planar texts. The scene texts possess numerous perspective distortions. They may be aligned in any direction with geometric distortions

(b) **Inter-character distance:** Characters present in a text line have some uniform distance in between them.

(c) **Color:** In a simple image the characters in a text usually have the similar or same color. Connected component-based approach usually makes use of this property. But complex images or color documents usually contain text strings with two or more colors for effective visualization

(d) **Edge:** Most caption and scene texts are designed to be easily read, therefore strong edges are placed at the boundaries of text and background.

The remaining part of this paper is organized as follows. Section II describes the related works. Proposed system is explained in Section III and the experiments and results in section IV. Finally the conclusion is given in the Section V.

2. RELATED WORKS

The various methodologies identified with content extraction incorporates locale based strategies, surface based techniques. The locale based strategies are essentially partitioned into two classifications: edge based [1] and associated segment based techniques. J.Gllavataet.al [3] proposed an associated part [2] based methodology for the content extraction. It depends on shading decrease system and OCR is utilized for character acknowledgment. It will just distinguish content with even arrangement. Low quality pictures won't be handled precisely. Zhong et al. [4] utilized a CC-based technique, which uses shading decrease. In that they quantize the shading space by investigating the shading histogram created in the RGB shading space. This is done basically dependent on the presumption that the content locales for the most part bunch together in this shading space and they involve a noteworthy bit of a picture. Every content part present will experience sifting stage utilizing various heuristics, for example, region, spatial arrangement and breadth. The presentation of this framework was assessed with testing in CD pictures and other book spread pictures. Kim et al. [5] use change guide to recognize overlay content. They proposed a strategy for overlay content recognition and extraction from complex recordings. The recognition strategy depends on the perception with respect to the presence of transient shading between embedded content and its nearby foundation. The progress guide is produced first which depends on its logarithmical change of force and adjusted immersion. At that point Linked cover is produced to make associated segments for every hopeful locale and after that every one of these associated segments is reshaped to have smooth limits. Looks into dependent on Caption content identification are proposed in [9]. Xiaoqing Liu,et.al [6] proposed a technique dependent on the properties of edges. This strategy isn't touchy to picture

shading/power. Despite the fact that it can deal with both printed and record pictures adequately. Since it for the most part investigations messages as obstructs the little picture districts and stroke are misidentified as content in territories containing huge characters. The proposed system uses wavelet and different morphological operations in a different way compared to existing techniques for the identification of text part. So a better efficiency is obtained in the extraction of text. Additional features are also added in our system with the enhancement of technology.

3. PROPOSED SYSTEM

The proposed technique works dependent on the way that the writings present in pictures have some interesting highlights which incorporate the properties of edges. The engineering of the proposed framework is appeared in Fig.2. The expert presented framework is for the most part partitioned into three modules: Edge map age module, Text region division module, Text acknowledgment module. The information is given as picture to the framework and the yield acquired is as a content document.

Edge Map Generation Module

The information is passed to this module. The info picture can be grayscale or shading, packed or uncompressed group. This module contains various strides in it. Right off the bat, the info picture experiences a preprocessing stage.

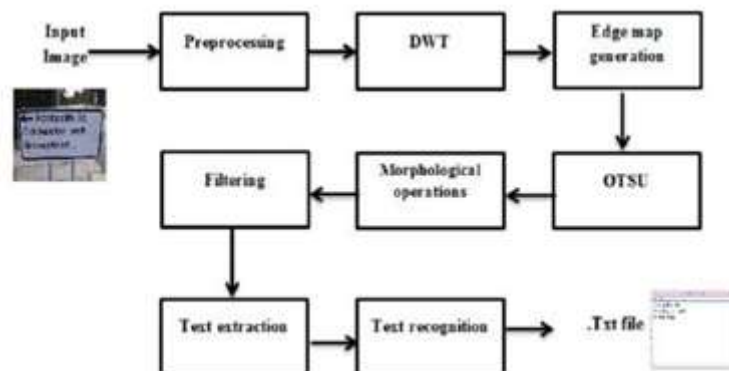


Fig.2. Architecture of the proposed system

Preprocessing

Edge quality and thickness are key attributes of that installed content in an intricate picture. This fills in as a crucial element for content extraction. In the expert presented calculation, the info taken is a shading or a dim scale picture. In the event that the info is a shading picture it experiences the preprocessing stage. In this stage, the shading picture is changed over to grayscale picture utilizing the condition beneath.

$$Y = 0.299R + 0.587G + 0.114B \quad (1)$$

The RGB picture is changed over to Hue-Saturation-Value (HSV) shading space. The Y of the above condition alludes to esteem part of the Hue-Saturation-Value (HSV) shading space. Along these lines RGB shading picture gets changed over to the dark scale picture. We can likewise process the picture by changing over it to YCbCr shading space and taking the Y, the luminance part for handling. On applying a middle sifting to this dark scale picture the clamor can be sifted through. The best known request measurements channel is the middle channel. It replaces the estimation of an alluded pix-el by another worth which is determined by taking the middle of the dark dimensions in the area of that pixel. Middle separating has incredible clamor decrease capacities with less obscuring contrasted with straight smoothing channels of comparable size. It is successful if there should be an occurrence of bipolar and unipolar rash clamor. After this separating step is done, the greater

part of clamor present in the picture gets evacuated while the edges in the picture will be as yet safeguarded. The Image Y is then handled with the 2-D discrete wavelet change.

Discrete Wavelet Transform

In this framework, the proposed calculation utilizes the Haar discrete wavelet trans-structure. The Haar DWT gives a well amazing asset which models a large portion of the qualities of the picture. The finished pictures are for the most part all around described by their edges. On applying Discrete Wavelet Transform (DWT), it is disintegrated into parts of recurrence area [8]. On applying the DWT the info picture is decayed to four sub-groups or parts. i.e one normal segment and three detail segments. To get the parts it needs to manage line and section bearing independently. First High Pass Filter (H.P.F) G and Low Pass Filter (L.P.F) H are abused for each column information and afterward are down inspected by 2 to get high and low recurrence segments of the line. Next the high and low pass channels are connected again for every high and low recurrence segments of the section and after that down inspected by 2. By method for the above handling, the four-sub band pictures are produced: HH1, HL1, LH1 and LL1

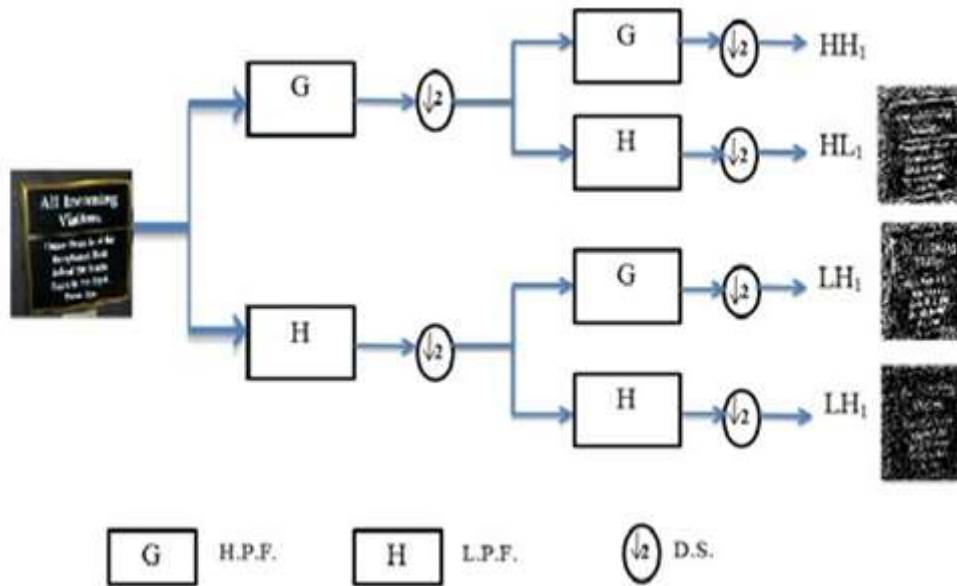


Fig.3. Block diagram of 2D DWT

The detail part sub-groups containing the vertical subtleties LH1, even subtleties HL1 and inclining subtleties LL1 are utilized to distinguish content edges present in the first picture. The way toward applying DWT to picture is appeared in Fig.3. The D.S. speaks to the down testing of the picture by 2.

Since sifting is done before applying DWT, the impact of commotion in the parts can be decreased. Presently edge location strategy is connected at every part. The wavelet work and the scaling capacity of Haar wavelets are characterized beneath.

$$\varphi(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ -\frac{1}{2}, & 1/2 \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

$$\varphi(t) = \begin{cases} 1, & 0 \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

The Haar is more straightforward than different wavelets which lessens the intricacy of the calculation. The upside of utilizing Haar wavelet is that it is the main wavelet that permits ideal limitation in the change space. Its coefficients are either 1 or - 1 and are genuine symmetric and symmetrical. On applying this, the segments with higher edge quality in indistinguishable ways can be discovered. From there on the limit is determined. This channels out the unfeasible edges. The second subordinate of force is utilized in the estimation of edge quality as it gives improved recognition of power places which prompts a portrayal of content in pictures that generally prompts a portrayal of content in pictures. The customary edge recognition channels may likewise have the option to give the comparable outcome yet it won't almost certainly identify three sorts of edges at any given moment. In this way, the handling time for the conventional edge discovery channels works slower than 2-D DWT.

Text Region Detection

The way toward distinguishing content areas can be part into two sub issues: identification and limitation. In the recognition step, general districts of the edge are delegated content or non-content. The size and state of these areas contrast in various calculation. For instance, a few calculations order 8x8 pixel squares, while others arrange individual output lines. In the limitation step, the consequences of location are assembled to frame at least one content occasions. This is typically spoken to as a jumping box around every content case. For the discovery of the content locale, the three definite sub segments which is acquired on applying Haar DWT is utilized. Sobel edge discovery calculation is connected to each point by point segment. In this way hc,vc,dc are pictures acquired after edge discovery in level ,vertical and corner to corner part. In this way hc contains every one of the edges flat way, vc contains every one of the edges vertical way and the dc contains every one of the edges askew way. Presently an edge guide is produced utilizing hc, vc and dc. It is created by utilizing weighted OR activity. The condition (4) utilized for the age of edge guide is demonstrated as follows

$$I = (40 * hc + 70 * vc + 30 * dc) \quad (4)$$

The I is the image obtained after applying the edgemap. hc, vc and dc indicates the horizontal, vertical and diagonal sub component respectively after sobel edge detection algorithm is applied. The edge map of an example image is shown in Fig 4 .This algorithm uses the sobel edge detector as it is more efficient to locate the strong edge present. Now the image I shows edges with all the possible text regions.



Fig.4. Edgemap of the image obtained

Text area segmentation

After edge guide is gotten, the binarization of the picture is finished. The parallel type of the edge guide is gotten by the thresholding task. For binarization of the picture the Otsu calculation is utilized. The thresholding activity evacuates the non content areas recognized up until this point. The means engaged with figuring Otsu limit are

- Reshape the picture to 1 dimensional.
- Compute histogram and qualities at every power level
- Initialize a network with qualities from 0 to 255
- Step through every single imaginable edge greatest force
- Compute the mean, weight and the fluctuation for the closer view and foundation.
- Calculate weight of the frontal area * difference of the closer view + weight of the foundation * fluctuation of the foundation.

Find the base worth.

The base worth determined is taken as the sift old for the binarization procedure. The confinement procedure of the content includes further improvement the content areas by taking out non-content locales from the picture. One of the principle properties the content displays is that all characters present will seem near one another in the picture. Accordingly it can shape a bunch. By considering this property we utilize morphological activities. The conceivable content pixels can be grouped together by utilizing the enlargement activity, in this way killing pixels that are a long way from the competitor content districts are conceivable. Expansion can be characterized as an activity that extends or improves the district of enthusiasm, by the use of the auxiliary component of the required shape and size. Huge organizing component is utilized in the widening procedure with the goal that locales which falsehood near one another can be improved. To restrict the content part unmistakably we utilize the morphological activities.

The morphological activities incorporate the disintegration, widening with line and circle as the organizing component is finished. The division of the distinguished content parts are done in this division stage. For that the associated segments are marked and the network utilized here is 8. The arrangement of properties, shape and estimations of the associated parts are processed. The territory and jumping box shape estimations according to the necessity are just considered. The zone of the associated part is considered as scalar as it speaks to the quantity of pixels in that district. Jumping box is put to the associated segment recognized and it must be the littlest containing the required content district [11]. The width, tallness and the up-per left corner position are recognized. Another worth can be figured by duplicating tallness and width of jumping box. The proportion of

this new worth and territory is taken. On the off chance that the proportion is under 1.5, at that point the locales so got are considered as content districts. The resultant picture got after expansion task may even comprise of some non-content districts or any clamor which are should have been disposed of. To kill clamor masses present in the picture a territory sifting is finished. After that lone those locales in the last picture whose zone is more noteworthy than or equivalent to 1/15 of the most extreme zone area recognized are held.

The Fig 5(a) demonstrates the picture after morphological activities and Fig 5(b) gives the picture after content extraction.



Fig.5. (a) Image after morphological operations, (b) Output image after segmentation.

Text Recognition

The picture got from the past stage is considered as contribution to this stage. Content acknowledgment is finished by utilizing the tesseract OCR. The twofold picture with polygonal content areas characterized is given as contribution to the tesseract. The preparing of the framework pursues a conventional pipe-line. The initial step is investigation of associated part. It diagrams the parts put away. This is a computationally costly structure choice, yet it has a critical favorable position. At this stage, diagrams are assembled, by settling, into Blobs. The Blobs subsequently got are sorted out into content lines. At that point the lines and the locales distinguished are dissected for getting fixed pitch or relative content. Presently the acquired content lines are isolated into words as per the character dispersing. Fixed pitch content is then cleaved quickly by utilizing the character cells. Relative content after that is separated into words by utilizing unmistakable spaces. Acknowledgment process here experiences a two-pass process. In the principal pass, an endeavor is made to perceive each word. Each word that is palatable perceived is then passed to a versatile classifier as a preparation information. The versatile classifier at that point will all the more precisely perceive message in the page. With the fruition of first pass, a subsequent disregard is then executed the page. In this the words that were not perceived in the past pass all around ok are perceived once more. A last stage resolves spaces, and finds the little content by checking the elective speculations for the x-tallness to it. After acknowledgment the UTF-8 code of the character are returned. This can be effectively changed over relating characters and are shown as yield as a content document.

4. EXPERIMENTS AND RESULTS

The framework is tried different things with enormous number of pictures and the outcomes got are portrayed here. It is obvious from the Table 1 that the proposed strategies have better accuracy rate and review rate contrasted with other existing procedures. The strategy proposed by Samarabandhu et.al demonstrates 91.8 % of accuracy and 96.6% of review while the system given by j.yanget. al have lesser rate of about 84.90% and 90.0 % accuracy and review. Be that as it may, our technique indicates 99.65% of precision rate and 99.8% of review. The other existing techniques proposed by Kim et.al and J.Gllavata shows much littler rate as given in the Table 1.

The exactness of the calculation in the proposed strategy is figured by checking the quantity of effectively found characters, which is taken as the ground truth. The exactness and review rate are determined by the condition (5) and (6)

$$\text{Precision rate} = \frac{\text{Correctly located}}{\text{Correctly located} + \text{false negative}} \times 100 \quad (5)$$

$$\text{Recall rate} = \frac{\text{Correctly located}}{\text{Correctly located} - \text{false negative}} \times 100 \quad (6)$$

Examinations with some current strategies are appeared table 1 which demonstrates an unmistakable improvement over existing techniques. The exhibitions of different techniques indicated are referred to from the distributed works. The proposed framework is executed utilizing Matlab[13].

Table 1 Comparisons with existing methods

Method	Precision rate (%)	Recall rate (%)
Proposed method	99.65	99.8
Samarabandhu et. al[6]	91.8	96.6
J. yang et. al[11]	84.90	90.0
K.C.kim et al[10]	63.7	82.8
J.Gllavata et.al[3]	83.9	88.7



Fig.6. The GUI showing the input image and the image with the text portion

5. CONCLUSION

In this paper, a generally straightforward, quick and viable technique for content discovery and extraction are proposed. The strategy utilizes DWT for the proficient working of the calculation. This procedure requires less handling time which is for the most part basic for constant applications and shows high exactness rate. A large portion of strategies bomb when the characters are not adjusted well or when the characters are excessively little. Those strategies likewise result in some of missing characters when the characters have exceptionally poor diverge from regard to the foundation. Be that as it may, the professional presented strategy isn't touchy to shading or power of picture, and furthermore the uneven brightening and reflection impacts.

This can be utilized in huge assortment of use fields, for example, vehicle tag recognition, object recognizable proof, distinguishing proof of different parts in modern robotization, portable robot route which identifies content based land marks, examination of specialized papers with the assistance of maps, outlines, and electric circuits and so on. This calculation is great at taking care of both scene content pictures and reports pictures adequately. Despite the fact that there are enormous quantities of calculations around there, it is seen that there is no single bound together methodology or calculation that fits for every one of the applications.

REFERENCES

- [1] Xin Zhang, Fuchun Sun, Lei Gu "A Combined AI-gorithm for Video Text Extraction" Seventh International Conference on Fuzzy Systems and Knowledge Discovery, 2010.
- [2] Ohya, A. Shio, and S. Akamatsu, "Recognizing Characters in Scene Images", IEEE Transactions on

- Pat-tern Analysis and Machine Intelligence, 1994, 214-224.
- [3] Gllavata, R. Ewerth, and B. Freisleben, "A robust algorithm for text detection in images" Proceedings of the 3rd International Symposium on Image and Signal Pro-cessing and Analysis, pp.611– 616, ISPA, 2003.
 - [4] Zhong, Yu.,Karu, K., and Jain, A.K." Locating text in complex color images" Proceedings of the Third Inter-national Conference on Document Analysis and Recogni-tion1995, 1, 14-16: 146-149.
 - [5] WonjunKim,Changick Kim, "A New Approach for Overlay Text Detection and Extraction From Complex Video Scene," IEEE Transactions on Image Processing, V.18 , No.2, pp. 401 – 411, 2009.
 - [6] Xiaoqing Liu, JagathSamarabandu "Multiscale Edge-Based Text Extraction from Complex Images," International Conference on Multimedia and Expo, pp.1721-1724, 2006
 - [7] Xiao-Wei Zhang, Xiong-Bo Zheng, Zhi-Juan Weng, "Text Localization Algorithm Under Back-ground Image Using Wavelet Transforms", Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition, Hong Kong, pp.30- 31, Aug. 2008.
 - [8] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," IEEE Transactions on Systems, man and Cybernet, 1979.
 - [9] Debapratim Sarkar, Raghunath Ghosh , "A Bottom-Up Approach of Line Segmentation from HandwrittenText"2009.
 - [10] K.C. Kim, H.R. Byun, Y.J. Song, Y.W. Choi, S.Y. Chi,K.K. Kim and Y.K Chung, Scene Text Extraction in Natural Scene Images using Hierarchical FeatureCom-bining and verification, Proceedings of the 17International Conference on Pattern Recognition (ICPR'04),IEEE.
 - [11] J. Yang, J. Gao, Y. Zhang, X. Chen and A. Waibel, "AnAutomatic Sign Recognition and Translation Sys-tem",Proceedings of the Workshop on Perceptive User Inter-faces(PUI'01), 2001, pp. 1-8.
 - [12] A.K Jain,"Fundamentals of Digital Image Pro-cessing",Englewood cliff, NJ: Prentice Hall, 1989.