

## Text Extraction of Vehicle Number Plate and Document Images Using Discrete Wavelet Transform in MATLAB

Akash Goel<sup>1</sup>, Prof. Yogesh Kumar Sharma<sup>2</sup>

<sup>1</sup>M.Tech Scholar, Computer Science & Engineering, Bhagwant Institute of Technology, MuzaffarNagar India

<sup>2</sup>Associate Professor, Computer Science & Engineering, S. D. College of Engg. & Tech., MuzaffarNagar

**Abstract:** Text Extraction from colour images is a challenging task in computer vision. The concept of text extraction is derived from the vehicle plate recognition and their characters extractions individually. Some examples of the applications are automatic image indexing, visual impaired people assistance or optical character reading, keyword searching in a document image. The continuous research has been done majorly that focuses on the text detection and localization in images. The paper focuses on the colour images text extraction using Two Dimensional Discrete Wavelet Transform (2D-DWT). The transform used for DWT is HAAR DWT with mathematical morphological concept. The speed of Haar DWT operates the fastest among all wavelets because its coefficients are either 1 or -1. It is one of the reasons that Haar DWT is used to detect edges of candidate text regions. Image sub bands contain both text edges and non-text edges. The work is carried out in MATLAB 12.0 image processing tool and simulated results are tested on vehicle plate text extraction and the document images.

**Keywords:** Text Localization, Mathematical Morphology Discrete Wavelet Transform (DWT), HAAR Transform, Mathematical Morphology

### I. Introduction

Recent advancement and research areas of image processing have much interest in content retrieval and derived in the perceptual and semantic content. Human perceptual includes color, shape pixel intensity and texture and semantic includes objects, events, interrupts and their relations. Contents of an image are described using texts, which are also easily and clearly describe the feature of an image. Since the text and characters data can be embedded in an image. Up to now it has been extracted by two basic techniques. These techniques are edge and connected component based technique. A text extraction system receives an input in the form of an image or a sequence of images. Text recognition and extraction problem can be divided into the following parts (i) Detection of text (ii) Localization of text, (iii) Tracking on text (iv) Extraction and enhancement of text, and (v) Recognition of text.

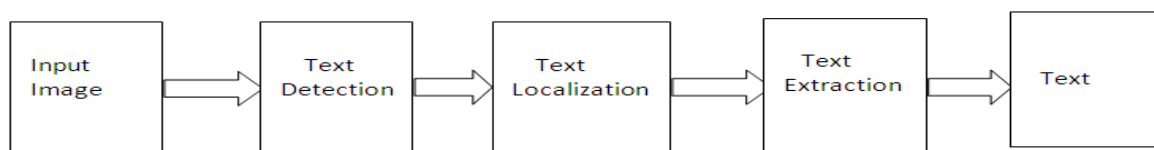


Fig. 1 Steps in text extraction

The steps of the text extraction are shown in figure1. The meaning of text detection is to detect the texts, which are presences in image. In this, threshold values are needed for scene based change detection because the portion occupied by a text region relative to the entire image is usually small. It is based on the difference between two consecutive frames and then used this scene change information for text detection. The methods of text localization are divided into two types: region and texture based. Regions defined methods use the properties of the color or gray scale in a text region or their differences with the corresponding properties of the background. Text extraction is done using two basic methods. First is region based while the other is based on texture.

#### 1.1 Region based methods

Region-based methods use the properties of the color or gray scale in a text region or their differences with the corresponding properties of the background. These methods also divided into two parts: connected component (cc) based and edge-based. CC-based methods use approach by grouping small components into successive very larger components until all regions are identified in the image. It s a four stages method, (i) binarization (ii) Tentative character component (iii) Character recognition (iv) Relaxation operation.

Edge based methods are based on contrast between the text and the background. Edges of the text boundary are identified and merged. Differential filters is applied to an input image and perform thresholding to

find vertical edges and smoothing operation that is used to eliminate small edges. The RGB components of a color input image are combining to give an intensity image Y as follows

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

Where, R, B and G are red, blue and green components of an image. Canny operators are used to detect edges in an image. One edge point in a small window is used in the estimation of scale and orientation to reduce the computational complexity and analysis. The different edges of the text are then enhanced using this scale information. Mathematical morphological dilation is performed to connect the edges into cluster.

### 1.2 Texture based methods

This methods use the observation that text in images have detect textural properties that distinguish them from the background. It is based on Wavelet Transform, FFT, spatial variance, etc. Which utilize a horizontal window to compute the spatial variance for pixels in a local neighborhood pixel, then the horizontal edges in an image are identified using a canny edge detector method, and the components of small edges are merged into longer lines. The research leads to focus on text extraction by edge based techniques.

The morphological operations like dilation [1] and erosions [2] 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, diagonal, horizontal 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. Text and character information extraction consists of 5 general steps which are detection, localization, tracking, extraction, enhancement, and recognition (OCR) [5]. In case of particular scene text focus is set on extraction. The step is done on located text area of image and its purpose was segmentation of characters from background that is separation of text pixels from background pixels. Images text extraction strongly affects recognition results and thus it is an important factor for good performance of the whole process. Text extraction methods are classified as threshold based and grouping based. Threshold based text extraction includes histogram based on Thresholding [8], adaptive or local Thresholding [9] and entropy based methods. Grouping based method encompasses clustering based, learning and region based methods. Clustering techniques are used to perform well on color text extraction [10]. Region based approaches, split and merge algorithm, including region growing and exploit spatial information to group character pixels more efficiently but drawback is dependence on parameter values of color images. Learning based methods mostly refer to multilayer perceptions and self organizing maps, but variation of scene text makes difficult to create representative training database.

## II. Mathematical Morphological Image Processing

Morphological image processing [9] is a set of non linear operations related to the shape or morphology of features in an image. Morphological operations [9, 10] do not depend on their numerical values but on the relative ordering of pixel values. The morphological operations are especially suited to the processing of binary images. An image is formed using small shape or template called a structuring element. The all possible locations in the image are positioned by structuring elements and it is compared with the corresponding neighbourhood of pixels. There are some operations to test whether the element "hits" or intersects the neighbourhood, while others test whether it "fits" the neighbourhood.

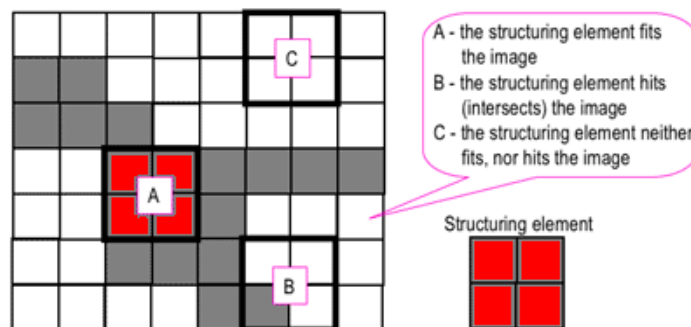


Fig. 2 An image with a structuring element

A new binary image is formed by morphological operation on a binary image in which the pixel has a non zero value at that location in the input image only if the test is successful. A structuring element is a small matrix of pixels with a value of zero and one. An image with structuring element is shown in figure 2. The size of structuring elements is specified by matrix dimensions. The shape of the structuring element is specified with the help of the patten of ones and zeros. The example of simple structuring element with one and zeros are shown in figure 3. Structuring element is a small part of the image and an origin of the structuring element is usually one of its pixels. In a structuring matrix the origin defines as the centre of the matrix.

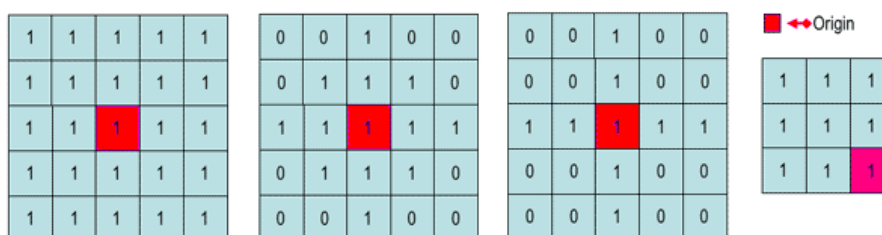


Fig.3 Examples of simple structuring elements

The structuring element is said to fit the image if each of its pixels is associated with the corresponding pixel of the neighbourhood for each of its pixels set to 1. Similarly, a structuring element is said to hit, or intersect, if at least for one of its pixels set to 1 the corresponding image pixel is also 1. Zero valued pixels of the structuring element indicate points where the corresponding image value is irrelevant and these values are ignored.

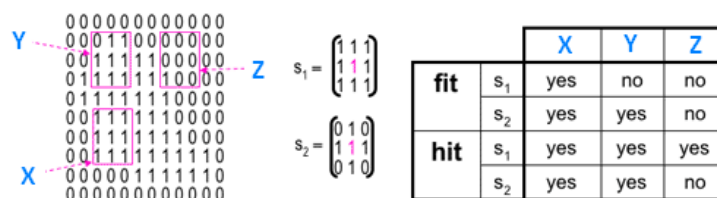


Fig. 4 Fitting and hitting of a binary image with structuring elements s1 and s2.

The basic morphological operations are dilation and erosion and expressed by a kernel operating of an input binary image. For a input binary image ‘X’, white pixels denote uniform regions and black pixels are denoting region boundaries. Dilation and Erosion work conceptually by translating a structuring element, over the image points. After it, intersection is done between the translated image coordinates and kernel coordinates.

### 2.1 Dilation

In dilation the image is expanded from its original shape and the way is determined with the help of structuring element. The size is choosing form structure element is generally 3 x 3 and it is smaller in size in the comparison to the image itself. In the dilation, the structuring element is reflected and shifted from left to right and from top to bottom. In each shift the process, overlapping is seen in similar pixels between the structuring element and of binary image. If there exists any overlapping then the pixels under the centre position of the structuring element will be turned to black or 1. For a reference image X and structuring element B, the dilation operation is given by

$$X \oplus B = \{z \left[ (\hat{B})_z \cap X \right] \subseteq X\} \quad (2)$$

Where  $\hat{B}$  is the image rotated about the origin? The equation (1) signifies that the image X is dilated by the structuring element B with an outcome Z. It would be that there will be at least one element in B that intersects with an element in X. In the case, the position where the structuring element is being centered on the image will be ‘ON. It is shown in figure in which black squares represents 1 and White Square represents 0.

In the example shown in figure 5, the centre of the structuring elements is at position 1 and no overlapping between black squares of X and black squares of B. That’s what square will remain same white at position 1 and structuring element will be shifted towards right. At position 2, it is seen that the black squares of B is overlapping or intersecting with the black square of X and the square will be changed to black. In the same way, the structuring element B is shifted from top to bottom and left to right on the image X to form the dilated image. The dilation process is similar to convolution process and expansion operators that enlarge binary

objects. It has many uses, but the major use is in bridging gaps in an image and extracting the features of an image.

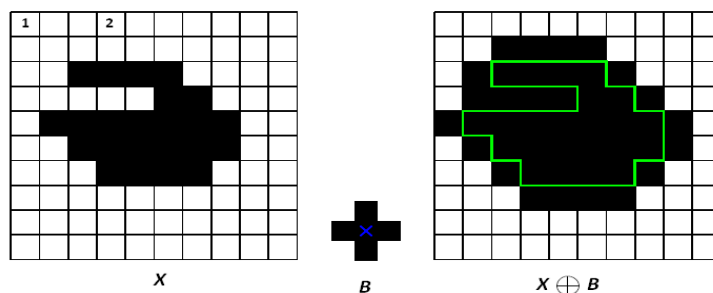


Fig. 5 Dilation of image X with structuring element B

### 2.2 Erosion

Erosion shrinks the image and it is the counter process of dilation because dilation enlarges the image. Structuring element is determining the way to shrink the image. Similarly, to the dilation process, in the erosion process structuring element will move from left to right and top to bottom. Center position of the structuring element will see whether there is a complete overlap with the structuring element or not. If there is no complete overlapping then the centre pixel will be set white or 0, indicated by the centre of the structuring element. The erosion process of image X with structuring element B is shown in figure 6.

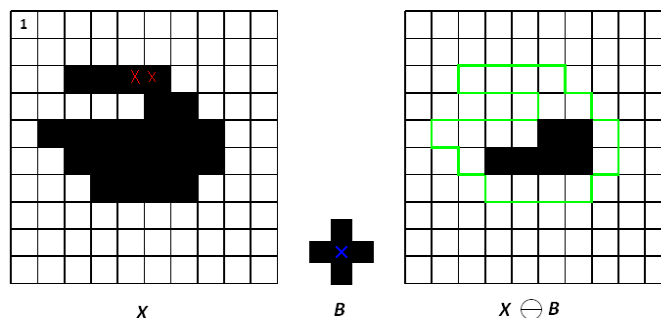


Fig.6 Erosion of image X with structuring element B

For a reference image X and structuring element B, the erosion operation is given by equation 2

$$X \ominus B = \left\{ z \left[ (\hat{B})_z \right] \subseteq X \right\} \quad (3)$$

The erosion process starts at position 1, and there is no complete overlapping. So, the pixels remains white at the position and structuring element is shifted to the right. At position 2, complete overlapping is not present, thus the black squares marked with \*\* will be turned to white. Erosion is the thinning operator to shrink the image. With the application of erosion, narrow regions can be eliminated while wider ones are thinned.

### 2.3 Opening

Opening is based on morphological operations dilation and erosion. Opening breaks narrow strips, smooths the inside of the narrow object contour and eliminates thin portions of the image. It is achieved by first applying erosion and then dilation operations on the image. Mathematically, it is expressed

$$X \circ B = (X \ominus B) \oplus B \quad (4)$$

Here X is an input image and b is the structuring element.

### 2.4 Closing

Closing operations is the opposite of the opening operation. In losing operation of image, the dilation operation is followed by an erosion operation. It smoothes contours and maintains shapes and sizes of the object. It protects coarse structures, closes small gaps and round off concave corners. Mathematically, closing process is expressed

$$X \cdot B = (X \ominus B) \oplus B \quad (5)$$

Here  $X$  is an input image and  $b$  is the structuring element.

The example in figure 7(a) shows the erosion of the dark blue squares by a disk, resulting in light green squares. Figure 7(b) shows the dilation of the dark blue square by a disk that result in the light green squares with rounded corners. Figure 7(c) shows the opening of the dark blue square by a disk resulting in the light green square with round corners. Figure 7(d) shows the closing of the dark blue square or the union of two squares by a disk resulting in the light green square with round corners.

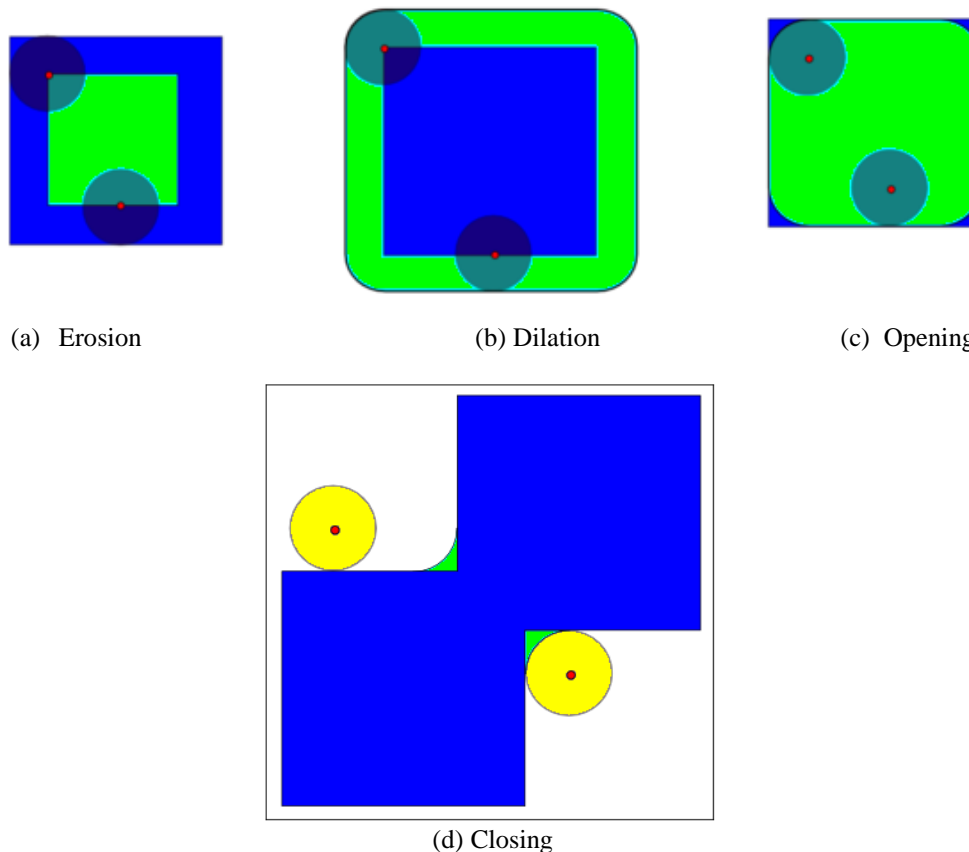


Fig. 7 Mathematical morphological operations

### III. Discrete Wavelet Transform (DWT)

A two dimensional digital image can be represented by a 2-D array  $X [m, n]$  with  $m$  rows and  $n$  columns, where  $m$  and  $n$  are positive integers of 2-D image. First process of a one dimensional DWT is performed on rows to get low frequency  $L$  and high frequency  $H$  components of the image. Again a one dimensional DWT is performed column wise on this intermediate result to form the final DWT coefficients such as  $LL, HL, LH, HH$ . They are called sub-bands.

The  $LL$  sub-band can be further decomposed into four sub-bands. This process can continue to the required number of levels. It is known multi level decomposition. The three level decomposition of the given digital image is as shown Fig. 8. High pass and low pass filters are used to decompose the image first row-wise and then column wise. So that similarly, inverse DWT is applied this is just opposite to the forward DWT, to get back the reconstructed image of the compression process. Various architectures have been proposed for computation of the DWT. These can be mainly classified as either convolution architectures or lifting based architectures. So that the number of computations required finding the DWT coefficients by the filter method is large for higher level of decomposition process. This leads to the implementation of new technique called lifting scheme for computing DWT coefficients. Because this scheme reduces the number of computations and also provides in-place computation of DWT coefficients

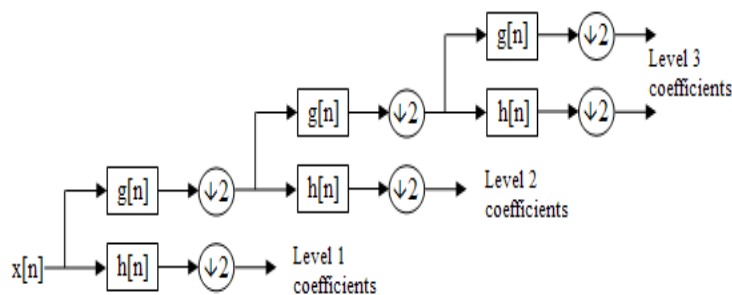
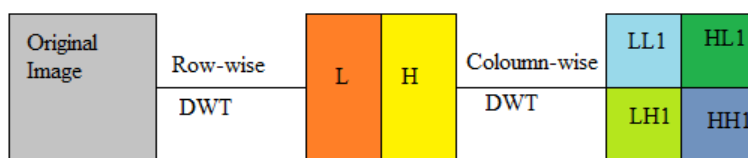
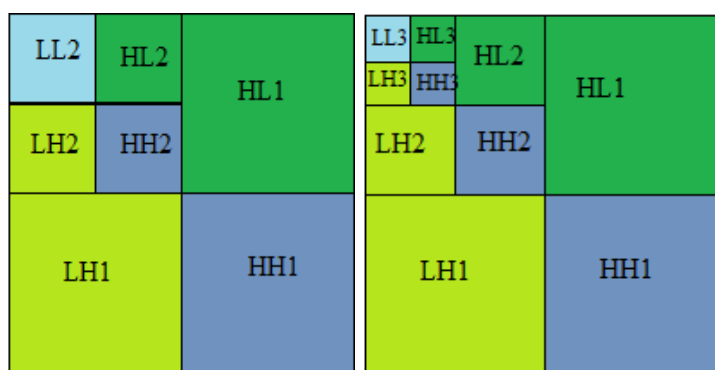


Fig 8 Two channel filter bank at level 3



(a) First level Decomposition of 2-D DWT



(b) Second level Decomposition (c) Third level Decomposition  
Fig.9 Different decomposition levels

So that in the discrete wavelet transform, the image signal can be analyzed by passing through an analysis filter bank followed by decimation operation. This analysis filter banks consist of a low-pass and high-pass filter at each decomposition stage of the process. When the signal passes through these filters such as Low-pass and High pass, it split through two bands. The low-pass filter of the filter bank, which corresponds to an averaging operation of the image sample, extracts the coarse information of the signal or image. The high-pass filter performed corresponds to a differencing operation, and extracts the detail information of the signal or image. Then output of the filtering operation is decimated by two. The two-dimensional transformation is accomplished by performing two separate one-dimensional transforms. The first, image is filtered along the row and decimated by two. Then it is followed by filtering the sub-bands image along the column and decimated by two. So this operation splits the image into four bands, such as, LL, LH, HL, and HH respectively. The different decomposition levels are shown in figure 9 (a), (b) and c. Two-dimensional discrete wavelet transform (2D DWT) decomposes an input image into four sub-bands, one average component (LL) and three detail components (LH, HL, HH) as shown in Fig. 10. In image processing, the multi-resolution of 2D DWT has been employed to detect edges of an original image. However, 2D DWT can detect three kinds of edges at a time while traditional edge detection filters cannot.

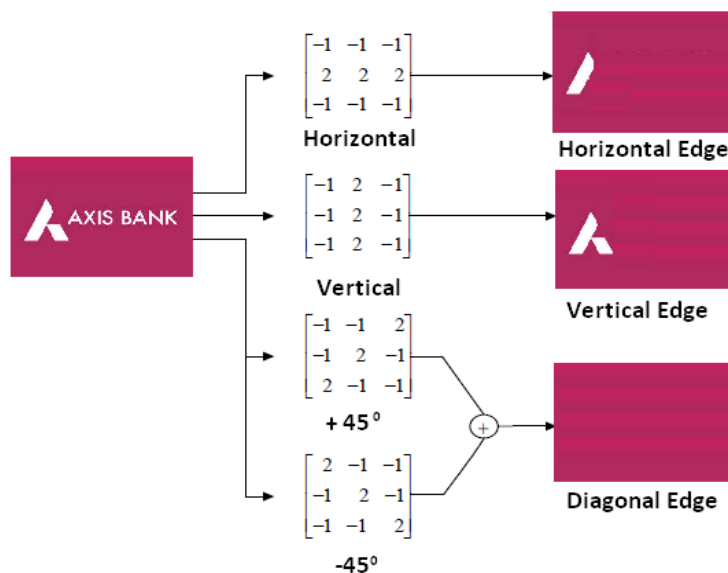


Fig 10 Detailed components

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 is slower than 2-D DWT. 9-7 taps DWT filters decompose this gray image into four sub-bands as shown in Fig. 10. As we can see, three kinds of edges present in the detail component sub-bands but look unobvious. The detected edges in image become more obvious and the processing time decreases, if we replace the 9-7 taps DWT filters with Haar DWT. The operation for Haar DWT is simpler than that of any other wavelets and applied to image processing especially in multi-resolution representation.

The morphological operations like erosion and dilations are used for better approach of refining text region extraction. Morphological operations are helpful in the removal of no texted regions. Various types of boundaries like horizontal, vertical, diagonal etc are clubbed together when they are segregated separately in unwanted non-text regions. 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 are associated with one other in diversified directions and are normally short. The Text and character extraction deploy both dilation and erosion for associating separated candidate text boundaries in every detail constituent sub band of the binary image.

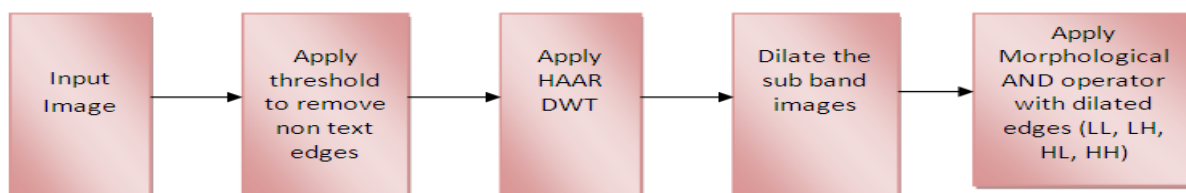


Fig.11 Steps of morphological AND operation

Fig. 11 matrix shows the example of a  $4 \times 4$  colour image. The wavelet coefficients can be obtained in gray-level image using addition and subtraction. 2-D DWT is achieved by two ordered 1-D DWT operations based on row and column values. First of all, row operation is performed to obtain the resultant matrix. Then it is transformed by the column operation and the final resulted 2-D Haar DWT matrix. 2-D Haar DWT decomposes [9] a gray-level image into one average component sub-band and three detail component sub-bands.

$$\begin{bmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & o & p \end{bmatrix} \quad \begin{bmatrix} (a+b) & (c+d) & (a-b) & (c-d) \\ (e+f) & (g+h) & (e-f) & (g-h) \\ (i+j) & (k+l) & (i-j) & (k-l) \\ (m+n) & (o+p) & (m-n) & (o-p) \end{bmatrix}$$

$$\begin{bmatrix} (a+b) + (e+f) & (c+d) + (g+h) & (a-b) + (e-f) & (c-d) + (g-h) \\ (i+j) + (m+n) & (k+l) + (o+p) & (i-j) + (m-n) & (k-l) + (o-p) \\ (a+b) - (e+f) & (c+d) - (g+h) & (a-b) - (e-f) & (c-d) - (g-h) \\ (i+j) - (m+n) & (k+l) - (o+p) & (i-j) - (m-n) & (k-l) - (o-p) \end{bmatrix}$$

Fig. 12 DWT matrix

#### IV. Results & Discussions

The simulation work of MATLAB text extraction is tested on different jpg and bmp colour images and the text extraction is done successfully. MATLAB tool 12.0 version is used and Image Processing tool is used for performance and simulation. The work is done on Windows XP. Figure 13 and (14) shows the MATLAB simulated results of the text extracted of car vehicle plate number.

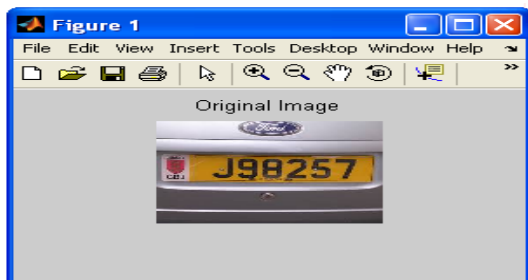
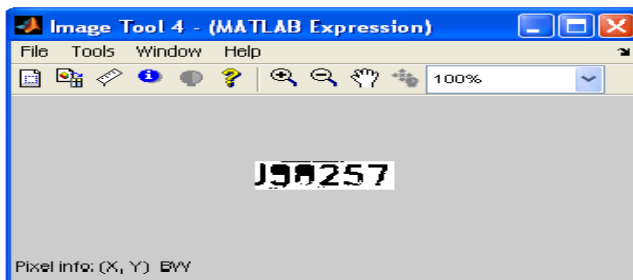


Fig. 13(a) Image 1 before text extraction



(b) text extracted image

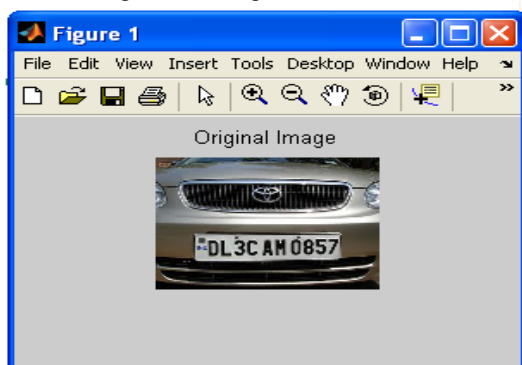


Fig. 14(a) Image 1 before text extraction



(b) text extracted image

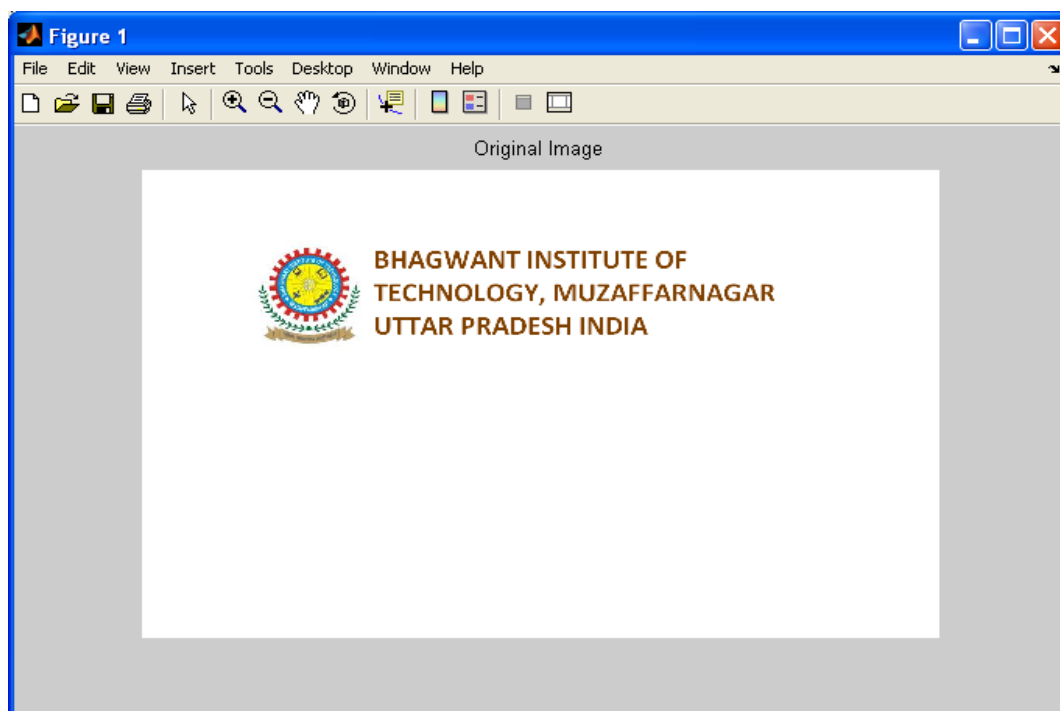


Fig. 15 (a) Original image before extraction

In the figure 13 (a) Original image is of vehicle plate which has the number J98257. After the MATLAB simulation the vehicle plate from the car body is extracted. In the similar way, figure 14(a) shows the



car plate number DL 3C AM 0857, after the simulation, the same number is extracted. Figure 15(a) and 16 (b) are the document images, which have horizontal text and extracted after MATLAB simulation.

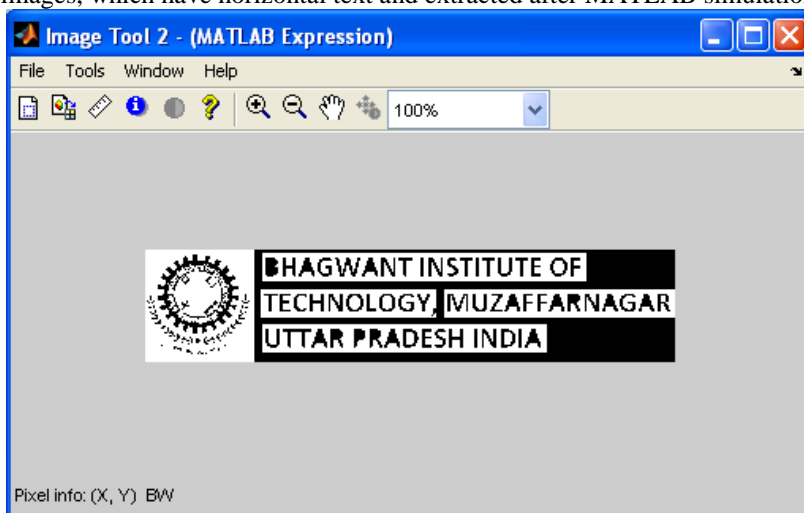


Fig. 15 (b) text extracted image

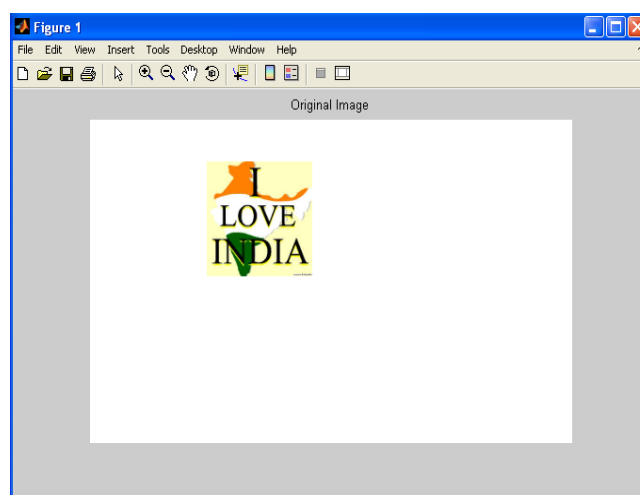


Fig. 16(a) Original image



Fig. 16 (b) text extracted image

## V. Conclusion

HAAR DWT has introduced a novel process of text extraction considering multiple cases of image with its textual contents. It also deploys methodology of sliding window for reading sub bands of high frequency. Morphological operations like dilation, erosion, opening and closing have been introduced to refine the text and non-text region appropriately. The MATLAB simulation is carried out on the different images with single / multiple text, multiple text of different sizes, images with uniform and non uniform images. This setup can be used to recover textual information from surveillance footage, satellite imaging, tollbooth as in a hybrid

approach is used to extract textual information from a video scene. In future the research work can be carried out on the large sample of document images and find out the particular text and extraction of special characters from a document image.

### References

- [1] A. Tognetti, F. Lorussi, R. Bartalesi, S. Quaglioni, M. Tesconi, G. Zupone, and D. De Rossi, "Wearable kinesthetic system for capturing and classifying upper limb gesture in post-stroke rehabilitation," *J. Neuroeng. Rehabil.*, vol. 2, no. 1, p. 8, Mar. 2005.
- [2] Chen, Datong, Bourlard, H., and Thiran J. P., 2001 "Text identification in complex background using SVM". IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings of the 2001, 2, 8-14:621-626.
- [3] Julinda Gillavata, Ralph Ewerth and Bernd Freisleben, A Robust algorithm for Text detection in images, Proceedings of the 3<sup>rd</sup> international symposium on Image and Signal Processing and Analysis, 2003.
- [4] J.D. Foley, A. van Dam, S.K. Feiner and J.F. Hughes, *Computer Graphics, Principles and Practice*, Addison-Wesley, Reading, 1990.
- [5] Keechul Jung, KwangIn Kim and Anil K. Jain, Text information extraction in images and video: a survey, The journal of the Pattern Recognition society, 2004.
- [6] Kongqiao Wang and Jari A. Kangas, Character location in scene images from digital camera, The journal of the Pattern Recognition society, March 2003.
- [7] 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 Feature Combining and verification" Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04), IEEE.
- [8] Lam, S. W., Wang, D., and Srihari, S. N., 1990. "Reading newspaper text" International Conference on Pattern Recognition Proceedings, 10th. I, 16-21: 703-705.
- [9] Mansi Agarwal, Adesh Kumar, Vimal Gupta "Text Extraction of Colour Images using Mathematical Morphology & HAAR Transform" IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 14, Issue 5 (Sep. - Oct. 2013), PP 78-87
- [10] Park, C. J., Moon, K. A., Oh, Weon- Geun, and Choi, H. M. 2000. An efficient of character string positions using morphological operator. IEEE International Conference on Systems, Man, and Cybernetics, 3, 8-11: 1616-1620.
- [11] Qixiang Ye, Qingming Huang, Wen Gao and Debin Zhao, Fast and Robust text detection in images and video frames, Image and Vision Computing 23, 2005.
- [12] Rainer Lienhart and Axel Wernicke, Localizing and Segmenting Text in Images and Videos, IEEE Transactions on Circuits and Systems for Video Technology, Vol.12, No.4, April 2002.
- [13] R.C. Gonzales and R.E. Woods, "Digital Image Processing", Addison-Wesley, Reading, 1992.
- [14] Victor Wu, Raghavan Manmatha, and Edward M. Riseman, Text Finder: An Automatic System to Detect and Recognize Text in Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 11, November 1999.
- [15] Xiaoqing Liu and JagathSamarabandu, A Simple and Fast Text Localization Algorithm for Indoor Mobile Robot Navigation, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5672, 2005.
- [16] Xiaoqing Liu and JagathSamarabandu, An Edge-based text region extraction algorithm for Indoor mobile robot navigation, Proceedings of the IEEE, July 2005.
- [17] Xiaoqing Liu and Jagath Samarabandu, Multiscale edge-based Text extraction from Complex images, IEEE, 2006.
- [18] Zhong, Yu., Karu, K., and Jain, A.K. 1995. Locating text in complex color images. Proceedings of the Third International Conference on Document Analysis and Recognition, 1, 14-16: 146-149.