DESIGN AND FPGA IMPLEMENTATION OF OPTICAL CHARACTER RECOGNIZATION METHOD USING VHDL

By
Pragyan Srivastava (R790211033)
Prakshi Rastogi (R790211034)

College of Engineering Studies
University of Petroleum & Energy Studies
Dehradun

April, 2015

DESIGN AND FPGA IMPLEMENTATION OF OPTICAL CHARACTER RECOGNIZATION METHOD USING VHDL

A project report submitted in partial fulfillment of the requirements for the Degree of
Bachelor of Technology
(Electronics Engineering)

By:
Pragyan Srivastava (R790211033)
Prakshi Rastogi (R790211034)

Under the guidance of
Dr. Adesh Kumar
Assistant Professor (Senior Grade)
UPES, Dehradun

Approved

………………………………..

Dean

College of Engineering Studies
University of Petroleum & Energy Studies
Dehradun

April, 2015

CERTIFICATE

This is to certify that the work contained in this report titled "Design and FPGA Of Optical Character Recognization Method Using VHDL" has been carried out by Pragyan Srivastava and Prakshi Rastogi under my supervision and has not been submitted elsewhere for a degree.

……………………………………….

Mentor
Dr. Adesh Kumar
Assistant Professor
(Senior Grade)
Deptt. Of Electronics, Instrumentation & Control Engineering
UPES, Dehradun

ACKNOWLEDGEMENT

We express our deepest thanks to **Dr Adesh Kumar**, our mentor for taking Part in useful decisions and giving necessary advices and guidance. We choose this moment to acknowledge his contribution gratefully.

It is our radiant sentiment to place on record deepest sense of gratitude to **Dr. Kamal Bansal (Dean, College of Engineering Studies)** and **Prof Sushabhan Choudhary (HOD, Deptt of EIC)** for careful and precious guidance which were extremely valuable for the project.

We have taken efforts in this project, however it would not have been possible without the kind support of many individuals. We would like to extend our sincere thanks to them.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

DWT	Discrete Wavelet Transform
OCR	Optical Character Recognization
DCT	Discrete Cosine Transform
AMA	Aligning & Merging Analysis
SVM	Support Vector Machine
IEEE	Institution of Electrical & Electronic Engineers
STFT	Short Time Fourier Transform
CWT	Continuous Wavelet Transform
VHDL	VHSI Hardware Description Language
VHSI	Very High Scale of Integration
FSM	Finite State Machine
RTL	Resistor Transistor Logic
GCLK	Gated Clock
FPGA	Field Programmable Gate Array
ADC	Analog to Digital Converter

DAC	Digital to Analog Converter
LUT	Look Up Tables
IOB	Input Output Buffer

ABSTRACT

In digital image processing, detection and extraction of text from a documentary image is found a challenging task, especially for inclined, vertical and circular text. In the extraction method, the image is segmental in text regions from a compound image. Digital image processing is an ever expanding and dynamic area with applications reaching out into our daily life such as digital signature, authentication, surveillance, medicine, space exploration, automated industry inspection and many others areas. These applications are involved in different processes like image enhancement, object detection, features extraction, colour imaging etc. Implementation of such applications on a general purpose computer can be easier, but every time it is not efficient due to additional constraints on memory and other peripheral devices. Out of the five senses – sight, hearing, touch, smell and taste, humans use to perceive their environment. Among all, sight of images is the most powerful. More than 99% of the activity of the human brain is involved in processing images from the visual cortex. A visual image is rich in information. A clustering based approach has been devised for estimating globally matched wavelet filters using a collection of ground truth images. Text extraction scheme can be extended for the segmentation of document images into text, picture that includes graphics background, tone images. There is an efficient yet simple method to extract text regions from video sequences or static images. The speed of Haar discrete wavelet transform (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 intensity of the text edges is also different from that of the non-text edges. Therefore, Thresholding is used to preliminary remove the non-text edges. Text regions of colour images are composed of horizontal edges, vertical edges and diagonal edges. Morphological dilation operators as AND, OR are applied to connect isolated text edges of each detail component sub-band in a transformed binary image. The simulation is carried out on MATLAB 2012 image processing tool.

CHAPTER 1
INTRODUCTION

1.1 REPRESENTATION OF 2D-IMAGE

A digital image is two dimensional discrete signals with N x N array of elements. Each element is the array is a number which represents the intensity of a sampled signal. The conversion of an image into digital is done using camera or scanner. Digital images can be read directly on a computer screen. For example a 4 x 4 image in matrix format and its 3D view is shown in figure 1.

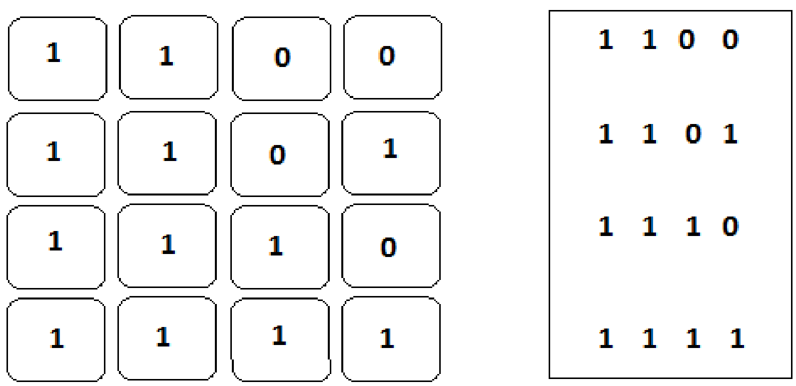


Fig. 1. Digital image representation

A pixel will have four neighbors exist in the EAST, WSET, NORTH and SOUTH direction. The four neighbors of the pixel P are shown in figure 2. A pixel P will have eight neighbors such as EAST, WEST, NORTH, SOUTH, NORTH-WEST (NW), NORTH-EAST (NE), SOUTH-EAST (SE) and SOUTH-WEST (SW). The eight pixels of the image are presented in figure 1.4.

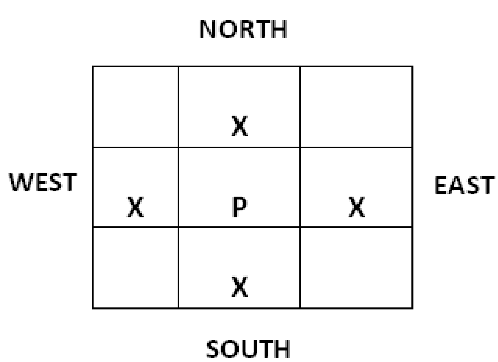


Fig. 2 Four Neighbors of pixel P

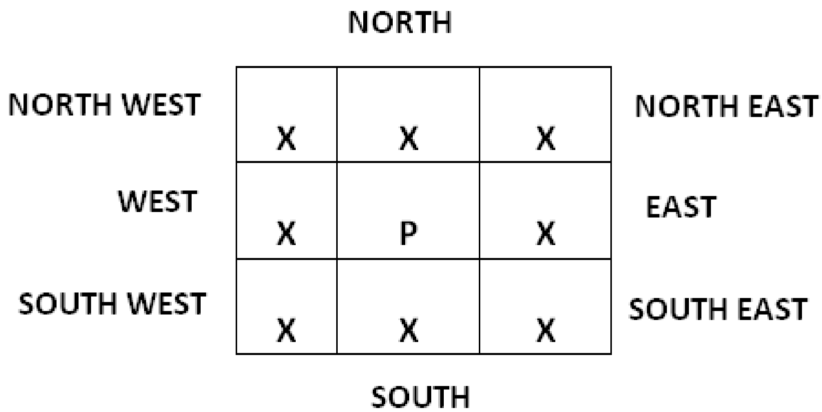


Fig. 3 Eight Neighbors of pixel P

1.2 INTRODUCTION TO IMAGE PROCESSING

Image processing is a rapidly growing field of digital signal processing [4]. Its growth has been analyzed by technological advances in computer vision, digital imaging, mass storage devices and computer processors [3] [17]. Analog systems follow the analog signal processing are replaced now by digital signal processing for their affordability and flexibility. Some examples of digital image processing are medical imaging, photography, [2] [18] security monitoring, remote sensing and video production. Such applications produce so volume of data every day that it is not possible to examine manually. Digital image processing techniques [1] [3] are concerned primarily with extracting important information from images.

Digital image processing can be implemented into digital chips. For example digital cameras [19] generally use dedicated digital image processing chips which are used to convert the raw data taken from image sensor into a colour image in a standard image file format [19]. Further these images are used in digital cameras to improve their quality. A software program is used for the modification in the image and can manipulate the images in different ways. Digital camera enable of viewing the histograms of images by which a photographer can understand rendered brightness range of each image shot more readily. Digital images play a very important role in our daily life applications such as magnetic resonance imaging [25], satellite television, and computer tomography. An image is defined as an array, or a matrix, of square pixels arranged in rows and columns. These are also called picture elements. An image can be represented in 2D configuration for a 3D scene [11]. An object can be represented by its numerical value by an image. An image is said a 2D function that represents some characteristics such as intensity, colour and brightness of any scene. It can be defined as a two variable function f(x,y) projected in a plane where f (x, y) defines the light intensity at particular point.

Pictures [1] [12] are the most common and convenient means of conveying or transmitting information. A picture is worth a thousand words [12]. Pictures concisely convey information about positions, sizes and inter-relationships [1] between objects. Human recognize the images as object which are represented in spatial information [12] [14] that we can recognize as objects. Human innate visual and mental abilities [13] are good at deriving information from such images, because of 75% of the information received by human is in pictorial form [10]. A sensed image is typically composed of digital picture pixel elements [11] located at the intersection of each row i and column j in each K bands of imagery associated with each pixel is a number known as Digital Number (DN) [10] or Brightness Value (BV) [3] that depicts the average radiance of a relatively small area within a scene as shown in figure 1.1. A smaller number indicates low average radiance [17] [19] from the area and the high number is an indicator of high radiant properties of the area. The portion of this image effects the reproduction of details within the scene. If it is possible to reduce the size of pixels, more scene detail is presented in digital representation. Integrated image processing [1] [2] , capture, and communication power on a portable, compact, [28] hand-held device is attracting and has increased interest from digital processing researchers with a goal of applying a diverse collection of vision [21] tasks on the small hand-held device.

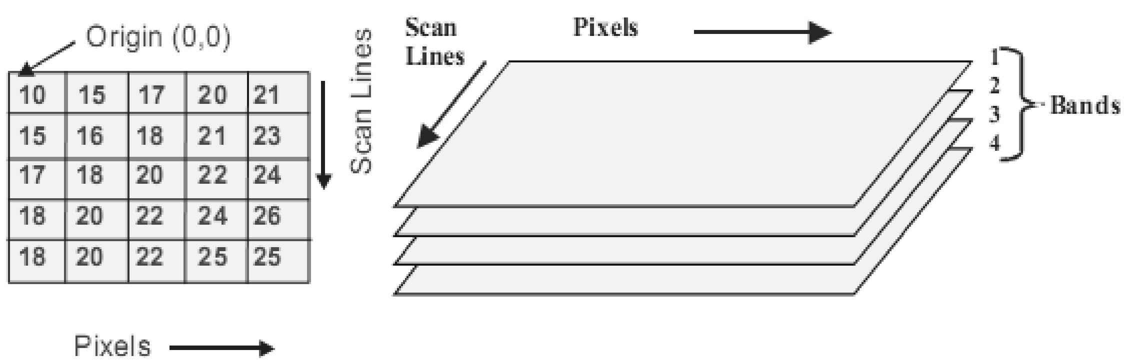


Fig. 4 Structure of a Digital Image and Multispectral Image

Research emphasizes on extract the text form the colored images using image processing tool. The extraction of text is carried out from any image using Haar Transform. Character reorganization from the image and character extraction form the colored image is also the objective of image extraction. A method to extract texts in images or video sequences using Haar discrete wavelet transform (Haar DWT) is also presented. The edges detection is accomplished by using 2-D Haar DWT and some of the non-text edges are removed using thresholding. After it, different morphological dilation operators to connect the isolated candidate text edges in each detail component sub-band of the binary image. Morphological dilation operators also extract the characters from the images using mathematical morphological and templates.

With the expansion of transmission technology over the last two decades, the demand for digital information increases dramatically. The advances in technology have created the utilization of digital image prevailing to an outsized extent. Still images are widely used in application like medical and satellite images. Digital image are comprised of an enormous amount of data. Reduction in the size of image data for both storing and transmission of digital images are becoming increasingly important as they find more application. Image extraction is a mapping from a higher dimensional space to a lower dimensional space. Image extraction play important role in many multimedia application, such as image storage and transmission. The basic goal of image extraction is to extract image data with minimum number of bits of an acceptable image quality. All image extraction algorithms strive to remove statistical redundancy and exploit perceptual irrelevancy while reducing the amount of data as much as possible. Data compression or image compression is the technique to reduce the redundancies in data representation in order to decrease data storage requirements and hence its coast of communication.

Decreasing the storage space requirement is equivalent to increasing the capacity of the storage medium and also increases communication bandwidth. Then the development of efficient compression techniques will continue to be a design challenge for future communication systems and advanced multimedia application. Image is represented by a combination of pixels and redundancy. A pixel is the portion of image that must be preserved permanently in its original form in order to correctly interpret the meaning or purpose of the image(data). Redundancy bits are the portion of image that can be removed when it is not needed or can be reinserted to interpret the image(data) when needed. Most often, the redundancy is reinserted in order to generate the original data (image) in its original form. The method to reduce the redundancy bits of image(data) is defined as image(data) compression. The redundancy bits in an image (data) representation are reduced such a way that it can be subsequently reinserted to recover the original image, and which is called decompression of the data. Integrated image capture, processing, and communication power on a compact, portable, hand-held device is attracting and has increased interest from computer vision researchers with a goal of applying a diverse collection of vision tasks on the small hand held device. The trend of vision applications is focused on camera phones and proposes solutions to the major technical challenges involving mobile vision and pattern recognition.

Recent studies in the field of computer vision and pattern recognition show a great amount of interest in content retrieval from images and videos. This content can be in the form of objects, texture, color, shape as well as the relationships between them. The semantic information provided by an image can be useful for content based image retrieval, and for indexing classification purposes. As stated by Jung, Kim and Jain in [4], text data is particularly interesting, because text can be used to easily and clearly describe the contents of an image. Text data can be embedded into an image or video in different font styles, sizes, colors, orientations and against a complex background is the problem of extracting the candidate text region becomes a challenging one. Also, current Optical Character Recognition (OCR) techniques can only handle text against a plain monochrome background and cannot extract text from a textured or complex background.

Different approaches for the extraction of text regions from images have been proposed based on basic properties of text. Text has some common distinctive characteristics in terms of frequency, spatial cohesion and orientation information. Spatial cohesion refers to the fact that text characters of the same string appear close to each other and are of similar height, orientation and spacing. Two of the main methods commonly used to determine spatial cohesion are based on edge [1] [2] and connected component [3] features of text characters. The fact that an image can be divided into categories depending on whether or not it contains any text data can also be used to classify candidate text regions. Thus other methods for text region detection, utilize classification techniques such as support vector machines, k-means clustering [7] and neural network based classifiers [10]. The algorithm proposed in [8] uses the focus of attention mechanism from visual perception to detect text regions. Text in images and video sequences pro-vide highly condensed information about the contents of the images or videos sequences and can be used for video browsing/retrieval in a large video database. Although texts provide important information about images or video sequences. Detection and segmentation of video or image sequences is not an easy problem. Text extraction is not easy for the following reasons. First of all, text sizes may vary from smaller to big and text fonts may vary in a wide range as well. Secondly, texts present in an image or a video sequence may have multiple colors and appear in very much cluttered background.

Many papers about the extraction of texts from static image or video sequence have been published in recent years. Those methods for texts extraction can be classified as either component based or texture based. In component based text extraction methods, different text regions are detected by analyzing the edges of the candidate regions or homogenous color/grayscale components that contain the characters. The logical AND operator is performed on dilated vertical edges and dilated horizontal edges to obtain candidate text regions. Real Text regions are then identified using support vector machine. Text regions usually have a special texture because they consist of identical character components. These components also contrast the background and hence text regions have a periodic horizontal intensity variation due to the horizontal alignment of characters. As a result, text regions can be segmented using texture features.

Most of the text extraction methods were applied to uncompressed images. Few of them proposed to extract texts in the compressed version of images. Zhong et al. [6] extracted captions from the compressed videos (MPEG video and JPEG image) based on Discrete Cosine Transform (DCT). DCT detects edges in different directions from the candidate image. Edge regions containing texts are then detected using a threshold afterward. Acharyya et al. [7] segmented texts in the document images based on wavelet scale-space features. The method used the M-band wavelet which decomposes an image into some M×M band pass channels so as to detect the text regions easily. The intensity of the candidate text edges are used to recognize the real text regions in an M-band image.

1.3 METHODOLOGY

Discrete Wavelet Transform (DWT) ,the two-dimensional extension of DWT 4, 12 is essential for transformation of two-dimensional signals, such as a digital image. A two-dimensional digital signal can be represented by a two-dimensional array X[M,N] with M rows and N columns, where M and N are nonnegative integers of 2D image array. The simple approach for two dimensional implementation of the DWT is to perform the one-dimensional DWT row-wise to produce an intermediate result and then perform the same one-dimensional DWT column-wise on this intermediate result to produce the final result. This is possible because the two-dimensional scaling functions can be expressed as separable functions which is the product of two-dimensional scaling function such as $\Phi_2(x,y) = \Phi_1(x)\Phi_1(y)$. The same is true for the wavelet function $\Psi(x,y)$as well. Applying the one-dimensional transform in each row of image, two sub-bands are produced in each row. When the low-frequency sub-bands of all the rows (L) are put together, it looks like a thin version (of size $M \times \frac{N}{2}$ of the input signal). Similarly put together the high-frequency sub-bands of all the rows to produce the H sub-band of size $M = \frac{N}{2}$, which contains mainly the high-frequency information around discontinuities in the input signal as edges. So that, applying a one dimensional DWT column-wise on these intermediate results L and H sub-bands four sub-bands LL, LH, HL, and HH 4, 8 of size(M/2 *N/2) are generated. LL is a coarser version of the original input signal. LH, HL, and HH are the high frequency sub-band containing the detail information of the image. It is also possible to apply one-dimensional DWT columnwise first and then row-wise to achieve the same result. The multi-resolution decomposition approach 8 in the two-dimensional signal is demonstrated . After the first level of decomposition, it generates four sub-bands LL1, HL1, LH1, and HH1. Considering the input signal is an image, the LL1 sub-band can be considered as a 2:1 sub-sampled (both horizontally and vertically) version of image. The other three sub-bands HL1, LH1, and HH1 8, 12 contain higher frequency detail information. These spatially oriented (horizontal, vertical or diagonal) sub-bands mostly contain information of local discontinuities in the image and the bulk of the energy in each of these three sub-bands is concentrated in the Author name / Procedia Computer

Science 00 (2015) 000–000 5 vicinity of areas corresponding to edge activities in the original image. In discrete wavelet transform six major steps of image compression. These encoding processes are following step-by-step. The encoding and decoding scheme of the DWT is listed step by step applied for the division of sub bands of images.

Fig 5 Encoding & Decoding Process Flow

The steps of encoding8 process are given below:

Step1: first the original image passed through a combination of filter, such as low-pass and high pass filter. These filters are applying each row.

Step 2: Then output image of the Bothe low-pass and high-pass filter is L1 and H1, these are combining into .

Step 3: After the filtering the combine output of these filters are down sampled by the 5.

Step 4: Now, again T1 has been passed through high pass filter and low filter by applying on each column.

Step 5: Let suppose the output of the step 4 is L2 and H5. Then H2 and L2 combine into

$$\Gamma_3 = \begin{bmatrix} L_2 \\ H_2 \end{bmatrix}$$

Step 6: After the filtering the combine output of these filters are down sampled by the 5. This is compressed image of the processing. Decoding Process The decoding process 8 is not exact reverse of the encoding process.

The steps of the decoding process are following.

Step 1: First extract low pass filter image and high pass filter image from the compressed image. The low pass filter image is taking by upper half rectangle matrix and high pass filter image is taking by down half rectangle matrix.

Step 2: These images are up sampled by 5.

Step 3: Then the summation of both images take into one image is called .

Step 4: Then again extract low pass filter image and high pass filter image by simply dividing vertically part of the image. First half is low pass filtered image part. And second half is high pass filter image.

Step 5: After then above process take summation of both images. It is the output of reconstructed image.

So, in the DWT we get very high extraction ratio, and also lose minimum amount of information. But if more than one level then get more extraction ratio but the reconstructed image is not identical to original image. The discrete wavelet transform is a very useful tool for signal processing and image analysis especially in multi-resolution representation. In DWT method, signals are decomposed into different components in the frequency domain. 1D DWT decomposes an input sequence into two components the average component and the detail component by calculations with a low-pass filter and a high-pass filter

.

2D DWT decomposes an input image into four sub-bands, one average component (LL) and three detail components (LH, HL, HH). In image processing we see that the multiresolution format of 2D DWT has been employed to detect edges of an original image.

LL	HL
LH	HH

Fig. 5 DWT decomposition scheme [4, 12]

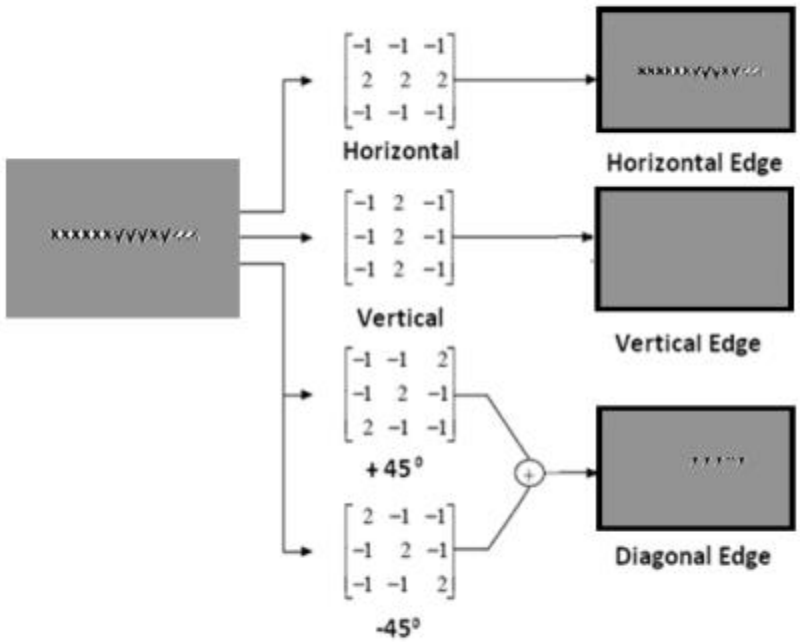


Fig 6 Edge Realization by DWT

In the process of wavelet transform, 2D DWT can detect three kinds of edges at a time while traditional edge detection filters. Three kinds of edges are detected using four kinds of mask operators with the help of traditional edge detection filters. That's what the processing times of the traditional edge detection filters is slower than 2-D DWT. DWT filters decompose the gray image into sub-bands as horizontal, vertical and diagonal edges 4 . In the filtration, three kinds of edges present in the detail component sub-bands but look unobvious. The processing time decreases with the replacement of the 9-7 taps DWT filters with Haar DWT, if the detected edges in image become more obvious. With the help of two ordered 1-D DWT, 2-D DWT is achieved by operations based on row and column values. Row line operation is performed to obtain the result and column operation is transformed by to get the final resulted 2-D Haar DWT. 2-D Haar DWT decomposes a gray-level image into one average component sub-band and three detail component sub-bands. The morphological operations like erosion and dilations are used for better approach of refining text region extraction

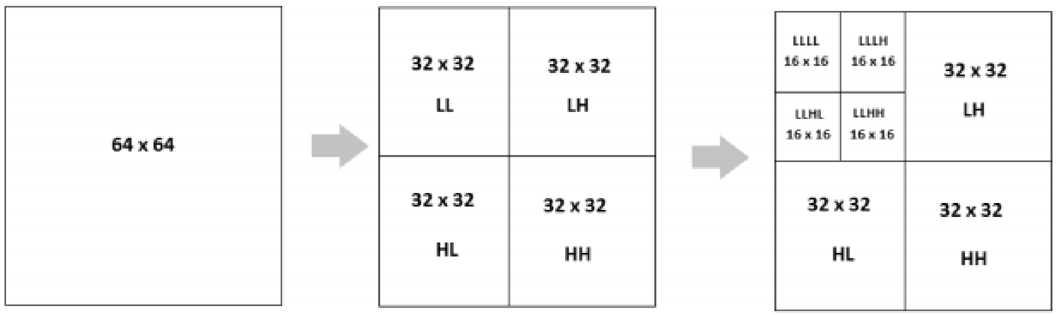


Fig. 7 Decomposition scheme of Haar Transform [8]

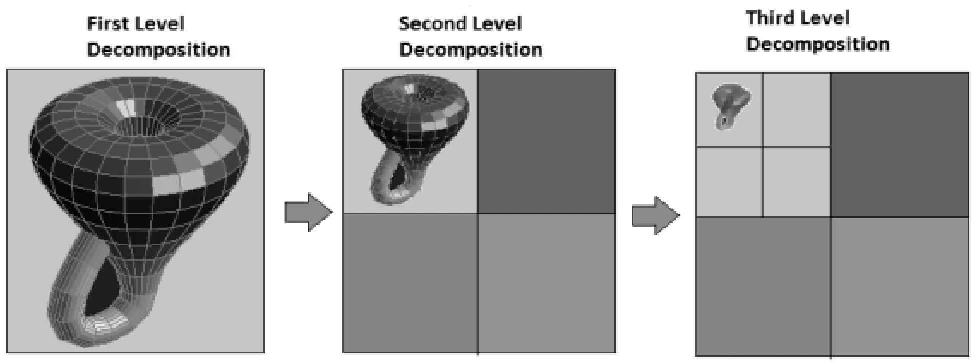


Fig 7 Decomposition of Image

Mathematical morphological operations are very much helpful in the removal of non-texted regions and various types of boundaries distribution as horizontal edge, vertical edge, diagonal etc are clubbed together when they are segregated separately in unwanted non-text regions. The text region of identified images consists of all these boundary and region information can be the area where such types of boundaries will be amalgamated. The edge and boundaries are associated with one other in diversified directions and are normally short. Dilation and erosion help for the text and character extraction deployment for associating separated candidate text boundaries in each detail constituent sub band of the binary image.

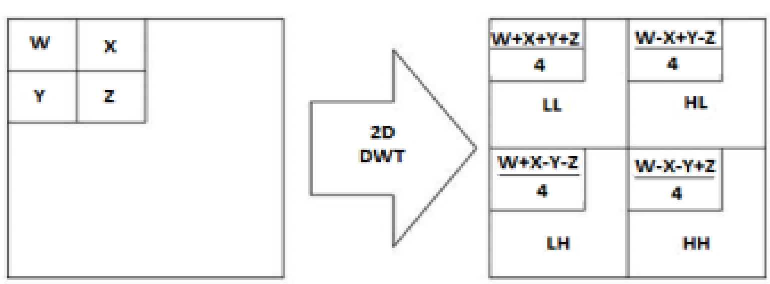


Fig 8 Division of Image into bands

1.4 OBJECTIVES

- Research objective is to extract the text form the coloured images using & the extraction of text is carried out from any image using Haar Transform.
- Character reorganization from the image and character extraction form the colored image is also the objective of image extraction.
- Use a method to extract texts in images or video sequences using Haar discrete wavelet transform (Haar DWT). The edges detection is accomplished by using 2-D Haar DWT and some of the non-text edges are removed using thresholding.
- After it, apply different morphological dilation operators to connect the isolated candidate text edges in each detail component sub-band of the binary image. Morphological dilation operators also extract the characters from the images using mathematical morphological and templates.

1.5 MOTIVATION

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 reorganization 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.

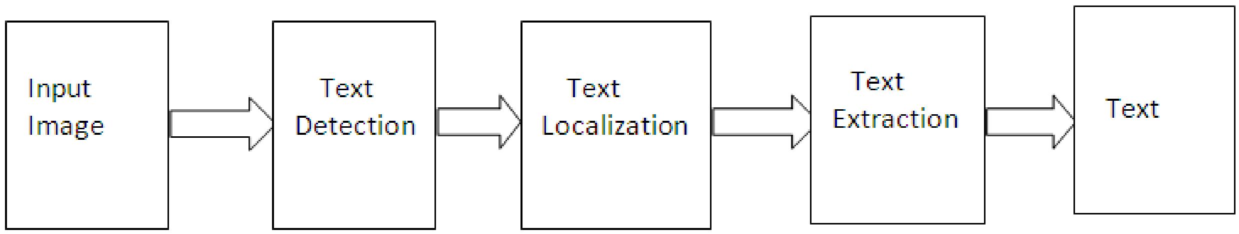


Fig. 9 Steps in text extraction

The meaning of text detection is to detect the text which is 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.

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 i n the image. It s a four-stage 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 R GB components of a color input image are combining to give an intensity image Y as follows:

$$Y = 0.299R + 0.587G + 0.114B$$

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. Texture based methods use the observation that text in images have detect textural properties that distinguish them from the background. It's 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 edgesare merged into longer lines. The research leads to focus on text extraction by edge based techniques.

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 colour 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.

CHAPTER 2
LITERATURE REVIEW

2.1 FINDINGS OF LITERATURE SURVEY

Various methods have been proposed in the past for detection and localization of text in images and videos. All these approaches are taken into consideration with different properties related to text in an image such as color, points, intensity, lines connected-components, edges etc. These properties of images are used to distinguish text regions from their background and/or other regions within the image. Wang and Kang as proposed the algorithm [5] which is based on color clustering. The binary input image is first pre-processed to remove any noise if present, and then the image is grouped into different color layers and a gray component. The approach helps in the utilization the fact that usually the color data in text characters is different from the color data in the background. Text regions on images are localized using connected component based heuristics from these layers. A method based on aligning and merging analysis (AMA) method is [6] used in which each row and column value is analyzed [5]. The analysis and experiments conducted show that the algorithm is robust in locating mostly Chinese and English characters in images, some false alarms occurred due to uneven lighting or reflection conditions in the test images

The text detection algorithm is also based on color continuity d intensity. With the addition of it, multi-resolution wavelet transforms combines low as well as high level image features for text region extraction. The algorithm for text finder is proposed in [7] is based on the orientation, frequency and spacing of text within an image. The segmentation used for texture based segmentation is used to distinguish text from its background. Furthermore, a bottom up process for chip generation is carried out which uses the spatial cohesion property of text characters. These chips are the collections of pixels in the image consisting of potential text strokes and edges. The experimental results show that the algorithm is robust in most cases except for very small text characters that are not properly detected. But, in the case of low contrast in the image, texture segmentation is suffered by misclassification which occurs in the different regions of text.

The approach used in [9], [11] utilizes a support vector machine (SVM) classifier to segment text from non-text in an image or video frame. Initially text is detected in multiscale images using edge based techniques, and projection profiles of the image [11] and morphological operations . These detected text regions are then verifiedusing wavelet features and SVM. This algorithm is robust with respect to variance in color and size of font as well as language. The goal of the research is to discover how the algorithms perform under variations of lighting, ,scale transformations and orientation of the text. The experimental results obtained were recorded based on criteria such as invariance with respect to lighting conditions, rotation, color, and distance from the camera (scale) as well as horizontal and/or vertical alignment of text in an image.

Several experiments have also been conducted for images containing 8 different font styles and text characters belonging to language types other than English. Also, the precision and recall rates (Equations (1) and (2)), have been computed based on the number of correctly detected words in an image in order to further evaluate the efficiency and robustness of each algorithm. The Precision rate is defined as the ratio of correctly detected words to the sum of correctly detected words plus false positives. The concept of false positives is derived from the regions in the image which are actually not characters of text regions, but the regions which have been detected by the algorithm as text regions

$$Precisionrate = \frac{Correctly\ detected\ words}{Correctly\ detected\ words + False\ Positives} \times 100\ \%$$

The Recall rate is defined as the ratio of correctly detected words to the sum of correctly detected words plus false negatives. The regions of the images are false regions in the image which are actually text characters regions, and these regions have not been detected by the algorithm:

$$Recall\ rate = \frac{Correctly\ detected\ words}{Correctly\ detected\ words + False\ Negatives} \times 100\ \%$$

Various papers present the status of research on text extraction.

2.2 DETAILED LITERATURE SURVEY

2.2.1 *Bin Cheng, Jianchao Yang, Student Member, IEEE, Shuicheng Yan, Senior Member, IEEE, Yun Fu, Member, IEEE, and Thomas S. Huang, Life Fellow, IEEE "2010 Learning With Graph for Image Analysis " IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 19, NO.4, APRIL (page 3,4)*

The graph construction procedure essentially determines the potential soft hose graph-oriented learning algorithm for image analysis. The main purpose of the paper is to propose a process to build the so-called directed graph, in which the vertices involve all the samples and the ingoing edge weights to each vertex describe it-norm driven reconstruction from the remaining samples and the noise. After it, a series of new algorithms for various matches in learning tasks, e.g., data clustering, sub space learning, and semi supervised learning, are derived upon the graphs. Compared with the conventional -nearest-neighbor graph and –ball graph the -graph possesses the advantages.

1) Greater robustness data noise,

2) Automatic sparsity, and

3) Adaptive neighborhood

2.2.2 *NamrataVaswani, YogeshRathi, Anthony Yezzi, and Allen Tannenbaum "Deform PF-MT Particle Filter With Mode Tracker for Tracking Nonaffine Contour Deformations" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 19 ,NO .4, APRIL 2010 (page 2-4)*

In this paper proposed algorithms are for tracking the boundary contour of a deforming object from an image sequence, which are nonaffine (local) deformation over consecutive frames is large and there is overlapping clutter, low contrast, occlusions, or outlier imagery. When there is the deformation of object in arbitrarily form each, or at least most, contour points can move independently. The contour deformation then forms an infinite (in practice point of view it is very large), dimensional space. Direct application of particle filters (PF) for large dimensional problems is impractically expensive. The property applied for it enables us to apply the particle filtering with mode tracking (PF-MT) idea that was proposed for such large dimensional problems in recent work. Most of the contour deformation is low spatial frequency we proposed to use the space of deformation at a sub sampled set of locations as the effective basis space. The algorithm based on it is called deform PF-MT.

2.2.3 *Srinivasa Reddy and B.N. Chatterji IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL.5, NO. 8, AUGUST 1996: An FFT-Based Technique for Translation, Rotation, and Scale-Invariant Image Registration B (page 2,3)*

Image registration is a fundamental task in image processing to overlay two or more images used. Image registration process can be loosely divided into the following classes: algorithms that use image pixel values directly, e.g., correlation methods .Algorithms that use the frequency domain, e.g., fast Fourier transform-based (FFT-based) methods .Algorithms that use low-level features such as edges and corners, e.g., feature-based methods and algorithms that use high-level features such as identified (parts objects, or relations between features, e.g., graph-theoretic methods. The registration method presented here uses the Fourier domain approach to match images that are translated, scaled, and rotated with respect to one another. Translation, scale, and rotation all have their counterpart in the Fourier domain. The methods of Fourier domain differ from other registration strategies because they search for the optimal match according to information in the frequency domain. In this correspondence, They presented an extension of the phase correlation technique for automatic image enhancement, which is characterized by its insensitivity to translation, scaling, rotation, and noise as well as by its low computational cost.

2.2.4 *ArnabSinha and Sumana Gupta "A Fast Nonparametric Non causal MRF-Based Texture Synthesis Scheme Using a Novel FKDE Algorithm" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL.19, NO.3, and MARCH 2010 (pg 2)*

This paper concludes that natural texture synthesis is an important problem in Image processing and computer vision. Synthesis models used for texture analysis can be broadly classified within two domains, spatial and transformed domains respectively. The spatial-domain models are also classified in two basic categories; Linear Model and non- linear models. All the models within image domain are sharing a common property, i.e., the pixel random variable is modeled as a function of neighborhood pixel random variables. In the transformed-domain, the synthesis algorithms can be classified into two categories, the first one is based upon the modeling of the transform domain with respect to a nonlinear model, and the second one is based upon modeling the transformed domain with respect to some nonlinear constraints. There use both spatial and transformed domain information.

2.2.5 *Yuichi Tanaka, Member, IEEE, Madoka Hasegawa, Member, IEEE, Shigeo Kato, Member, IEEE, MasaakiIkehara, Senior Member, IEEE, and Truong Q. Nguyen, Fellow, IEEE "Adaptive Directional Wavelet Transform Based on Directional Pre filtering" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 19, NO.4, APRIL 2010 (page 9)*

In this paper transform direction determining methods for the adaptive directional WT (Wavelet transforms) have been discussed. Wavelet Transforms are regarded as pre filtering steps by the 2-D filter bank or the directional WT for fixed directions. The obtained sub bands are used as reference frames for the direction calculation of the adaptive directional WT. The proposed framework requires fewer filtering operations than the conventional FS-WTs. In image and video processing using wavelet transform (WT), multi resolution decomposition is one of the most important features. It represents an image by several. Traditionally, 2-D WT is based on1-D filterings along horizontal and vertical directions. Four sub bands LL, LH, HL, and HH are obtained from the first level 2 DWT by extracting edges, i.e., regions containing high-frequency, along horizontal, vertical or diagonal orientation from the original image.

2.2.6 *Yuichi Tanaka, Madoka Hasegawa, Shigeo Kato, MasaakiIkehara and Truong Q. Nguyen, (2010) (page 3 and 6)*

In "Adaptive Directional Wavelet Transform Based on Directional Pre filtering" have proposed that the transform direction determining methods for the adaptive directional WT (Wavelet transforms) have been discussed. They are regarded as pre filtering steps by the 2-D filter bank or the directional WT for fixed directions. The obtained sub bands are used as reference frames for the direction calculation of the adaptive directional WT. The proposed framework requires fewer filtering operations than the conventional FS-WTs. In image and video processing using wavelet transform (WT), multi resolution decomposition is one of the most important features. It represents an image by several. Traditionally, 2-D WT is based on1-D filterings along horizontal and vertical directions. Four sub bands LL, LH, HL, and HH are obtained from the first level 2

DWT by extracting edges, i.e., regions containing high-frequency, along horizontal, vertical or diagonal orientation from the original image.

2.2.7 *Chung-Wei Liang and Po-Yueh Chen∗ DWT Based Text Localization International Journal of Applied Science and Engineering 2004. 2, 1: 105-116 (pg1)*

This paper presents an efficient yet simple method to extract text regions from static images or video sequences. Actual operation speed of Haar discrete wavelet transform (DWT) operates the fastest among all wavelets because its coefficients are either 1 or -1. This is one of the reasons we employ Haar DWT to detect edges of candidate text regions. Detailed sub bands as component contain both text edges and non-text edges. Although, the intensity of the text edges is different from that of the non-text edges, yet texture analysis is possible. Therefore, It is possible to apply thresholding to preliminary remove the non-text edges. The regions of texture are composed of vertical edges, horizontal edges and diagonal edges. Mathematical morphological dilation operators are applied to connect isolated text edges of each detail component sub-band in a transformed binary image. With the help of experiment results, the real components of text regions are the overlapped portion of three kinds of dilated edges. So, we can apply morphological AND operator to three kinds of dilated edges and obtain the final text regions correctly.

2.2.8 *Dr.N.Krishnan, C. Nelson Kennedy Babu S. Ravi and Josphine Thavamani SEGMENTATION OF TEXT FROM COMPOUND IMAGES, International conference of Computational intelligence and multimedia applications 2007(page 1-3*)

A method of text extraction from images is proposed using the Haar Discrete Wavelet Transform, the Sobel edge detector, the weighted OR operator, thresholding and the morphological dilation operator. These mathematical tools are integrated to detect the text regions from the complicated images. The proposed method is robust against language and font size of the texts. The proposed method is also used to decompose the blocks including multi-line texts into single line text. According to the experimental results, the proposed method is proved to be efficient for extracting the text regions from the images. Haar DWT detects three kinds of edges of the original image. 2D Haar DWT decomposes an input image into 4 sub-bands, one average component (LL) and 3 detail components (LH, HL, HH).The illumination components are transformed to the wavelet domain using Haar wavelet. This stage results are in the four bands LL, LH ,HL, and HH sub image coefficients.

2.2.9 *NidhiSethi, Ram Krishna Dehradun Institute of Technology, Dehradun Uttarakhand , India Image Compression Using Haar Wavelet Transform Computer Engineering and Intelligent Systems ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online)*

This paper reported is aimed at developing computationally efficient and effective algorithm for lossy image compression using wavelet techniques. The algorithm used in it or proposed algorithm is developed to compress the image in a time-efficient manner. The binary imaging results obtained concerning the reconstructed image quality as well as preservation of significant image details is promising. The encoding time is reduced with little degradation in image quality compared to existing methods. Haar Wavelet Transform has received a great amount of attention in the last decade for texture analysis. Wavelet based image compression introduces no blocky artifacts in the decompressed image. The Image which decompressed is much smoother and pleasant to eyes. It can be achieved much higher compression ratios much regardless of the amount of compression achieved. Main interesting feature of wavelet is that we can improve the quality of the image by adding detail information. The attractive feature for what is known as progressive transmission of images. They have chosen Haar Wavelet Transformation for image compression.

2.2.10 *Sunil Kumar, Rajat Gupta, Nitin Khanna, Student Member, IEEE, Santanu Chaudhury, and Shiv Dutt Joshi, Text Extraction and Document Image Segmentation Using Matched Wavelets and MRF Model "IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 16, NO. 8, AUGUST 2007"*

The paper proposed a novel scheme for the extraction of textual areas of an image using globally matched wavelet filters. A technique based on clustering approach has been devised for estimating globally matched wavelet filters using a collection of ground truth images. In the paper it is extended the text extraction scheme for the segmentation of document images into texture analysis, background, and picture components which include graphics and continuous tone images. In the method, multiple classifiers, two class fisher classifiers have been used for this purpose. It is also exploited contextual information by using a Markov random field formulation based pixel labeling scheme for refining of the segmentation results. Experimental results are carried out in MATLAB effectiveness of our approach. It has been presented a novel technique for locating the text part based on textural attributes using GMWs. The filtering and the feature extraction operations account for most of the required computations. Anyhow, their method is very simple, efficient and computationally less expensive and compared to other existing methods the dimensionality, and so, the computation of the feature space is considerably reduced.

CHAPTER 3
THEORETICAL DEVELOPMENT

3.1 INTRODUCTION TO WAVELET & DWT

The transform of an image signal is just another form of representing the image signal. The transform does not change the information content present in the signal. Wavelet Transform provides a time and frequency representation of the image signal. It was developed to overcome the short coming of the Short Time Fourier Transform (STFT), which can also be used to analyze non-stationary signals. Where STFT gives a constant resolution at all frequencies bands, the Wavelet Transform uses multi-resolution technique by which different frequencies are analyzed with different resolutions

A wave is an oscillating function of time or frequency and it is periodic. The wavelets are localized waves. It has their energy concentrated in time or frequency and they are suited to analysis of transient time signals. Where Fourier Transform and STFT use to analyze signals in the wave, the Wavelet Transform uses finite energy wavelets.

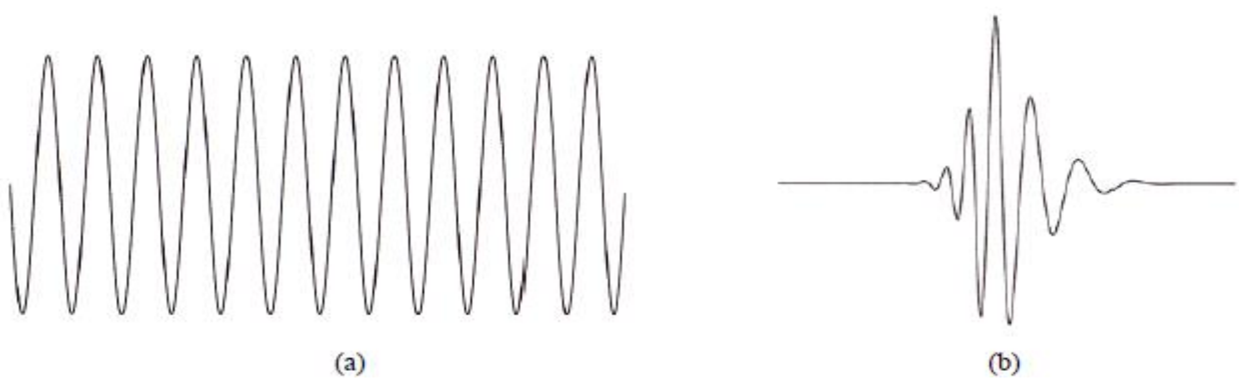


Fig. 10 Demonstrations of (a) a Wave and (b) a Wavelet [6]

The wavelet analysis is similar to the STFT analysis. In the wavelet signal to be analyzed is multiplied with a wavelet function same as signal is multiplied with a window function in Short Time Fourier Transform, and then the transform is computed for each segment generated signal of the wave. In the Wavelet Transform, the width of the wavelet function changes with every spectral component. At high frequencies the Wavelet Transform gives good time resolution and poor frequency resolution, and at low frequencies; Wavelet Transform gives good frequency resolution and poor time resolution. Mathematically a wave is expressed as an oscillating or sinusoidal function of time or frequency (space). Fourier transform expands an arbitrary signal in terms of infinite number of harmonics of its sinusoidal functions. Fourier representation of signals is known to be very effective in analysis of periodic signals. But in contrast to a sinusoidal function, wavelet is a small wave whose energy is concentrated in time domain. Properties of wavelets allow both frequency and time analysis of signals simultaneously because the energy of wavelets is concentrated in time and still possesses the wave-like (periodic) characteristics. The wavelet representation thus provides a versatile mathematical tool to analyses transient, non-stationary (time-variant) signals that may not be statistically predictable especially at the region of discontinuities – a special feature that is typical of images having discontinuities at the edges.

3.2 ORIGIN OF WAVELET TRANSFORM

Fourier transform is powerful tool that has been available to signal analysis for many years. It gives information regarding the frequency contains of a signal. However, the problem with using Fourier transforms is that frequency analysis cannot offer both goods frequency and time resolution at the same time. A Fourier transform does not give information about the time at which a particular frequency has occurred in the signal. Hence, a Fourier transform is not an effective tools to analyses a non-stationary signal. To overcome this problem, windowed Fourier transform, or short-time Fourier transform, was introduced. Even though a short-time Fourier transform has the ability to provide time information, multi-resolution is not possible with short-time Fourier transforms. Wavelet is the answers of the multi-resolution problem. A wavelet has the important property of not having a fixed-width sampling window. The wavelet transform is classified into two categories (i) Continuous wavelet transform, and (ii) Discrete wavelet transform. For long signal, continuous wavelet transform can be time consuming since it needs to integrate over all times. To overcome the time complexity, discrete wavelet transform was introduced. Discrete wavelet transforms can be implemented through sub-band coding. The DWT is useful in image processing because it can simultaneously localize signals in time and scale, whereas the DFT or DCT can localize only in the frequency domain.

3.3 CONTINUOUS WAVELET TRANSFORMS

The Continuous Wavelet Transform (CWT) is provided by ψ(t) isthe basis function or the mother wavelet. All the wavelet functions used in the transformation are derived from the mother wavelet through translation (shifting) and scaling (dilation or compression).Wavelets are functions generated from one single function (basis function) called the prototype or mother wavelet by dilations (scaling) and translations (shifts) in time (frequency) domain. The other main requirement is that the mother wavelet must have finite energy:

$$\int_{-\infty}^{+\infty} |\boldsymbol{\Psi}(\boldsymbol{t})|^{2} \boldsymbol{dt} < \infty$$

A mother wavelet and its scaled versions are depicted below indicating the effect of scaling. IfΨ(t) is denoted the mother wavelet function, the other wavelets Ψa, b (t) can be represented as:

$$\mathbf{\Psi a, b(t)} = \frac{\mathbf{1}}{\sqrt{|\mathbf{a}|}} \mathbf{\Psi}\left(\frac{\mathbf{t-b}}{\mathbf{a}}\right) \tag{3.1}$$

The variables *a*and*b* represent the parameters for dilations or scaling and translations or shifting respectively in the time axis.Where *a* and*b* are two arbitrary real numbers. From Eq. 3.1, it is obvious that the mother wavelet can be essentially represented as

$$\boldsymbol{\Psi}(\boldsymbol{t}) = \boldsymbol{\Psi 1}, \boldsymbol{0}(\boldsymbol{t}) \tag{3.2}$$

Where any arbitrary $a \neq 1$ and $b = 0$, it is possible to derive that

$$\mathbf{\Psi a, b(t)} = \frac{\mathbf{1}}{\sqrt{|\mathbf{a}|}} \mathbf{\Psi}\left(\frac{\mathbf{t-b}}{\mathbf{a}}\right) \tag{3.3}$$

As shown in Eq. 3.3, Ψa, 0(t) is nothing but a time-scaled (by *a*) and amplitude-scaled (by$\sqrt{|a|}$) version of the mother wavelet function Ψ(t) in Eq. 3.5. The parameter *a* causes contraction of $\Psi(\mathrm{t})$in the time axis when a < 1 and expression or stretching when *a* > 1. That's why the parameter *a* is called the dilation (scaling) parameter. For *a* < 0, the function Ψa,(t) results in time reversal with dilation or scaling. Mathematically, substituting *t* in Eq. 3.3 by *t-b* to cause a translation or shift in the time axis resulting in the wavelet function Ψa, *b* (t) as shown in Eq. 3.1. The function Ψa,(t) is a shift of Ψa,0(t) in right along the time axis by an amount *b* when *b*

> 0 whereas it is a shift in left along the time axis by an amount b when b < 0. That's why the variable b represents the translation in time (shift in frequency) domain.

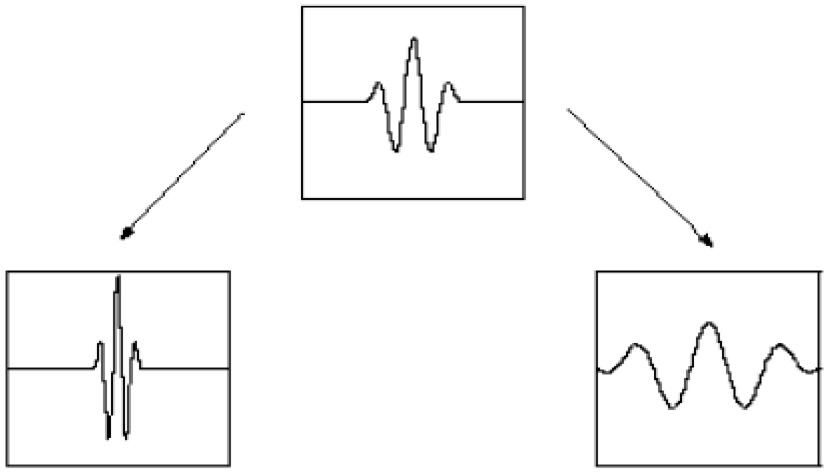


Fig. 11 Mother wavelet and its scaled versions [6]

Fig. 3.3 shows an illustration of a mother wavelet and its dilations in the time domain with the dilation parameter $a = \alpha$. For the mother wavelet Ψ t) shown in Fig.:3.3(a), a contraction of the signal in the time axis when $\alpha < 1$ is shown in Fig. 3.4(b) and expansion of the signal in the time axis when $\alpha > 1$ is shown in Fig.:3.3(c). Based on this definition of wavelets, the wavelet transform (WT) of a function (signal) f (t) is mathematically represented by

$$\boldsymbol{W(a,b) = \int_{-\infty}^{+\infty} \Psi a, b(t) f(t) dt} \tag{3.4}$$

The inverse transform can be reconstruct f(t) from W(a, b) is mathematically represented

$$\boldsymbol{f(t) = \frac{1}{c} \int_{a=-\infty}^{+\infty} \int_{b=-\infty}^{+\infty} \frac{1}{|a|^2} W(a,b) \Psi a, b(t) dadb} \tag{3.5}$$

Where

$$\boldsymbol{C = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega} \tag{3.6}$$

Ψ (ω) is the Fourier transform of the mother wavelet ψ (t).

C is required to be finite, which leads to one of the required properties of a mother wavelet. Since C must be finite, then $\Psi(0) = 0$ to avoid a singularity in the integral, and thus the $\Psi(t)$must have zero mean. This condition can be stated asIf a and$b$ are two continuous (non-discrete) variables and *f* (t) is also a continuous function, W (a,) is called the continuous wavelet transform (CWT). Hence the CWT maps a one-dimensional function *f*(t) to a function W(a, b) of two continuous real variables a (dilation) and b (translation).

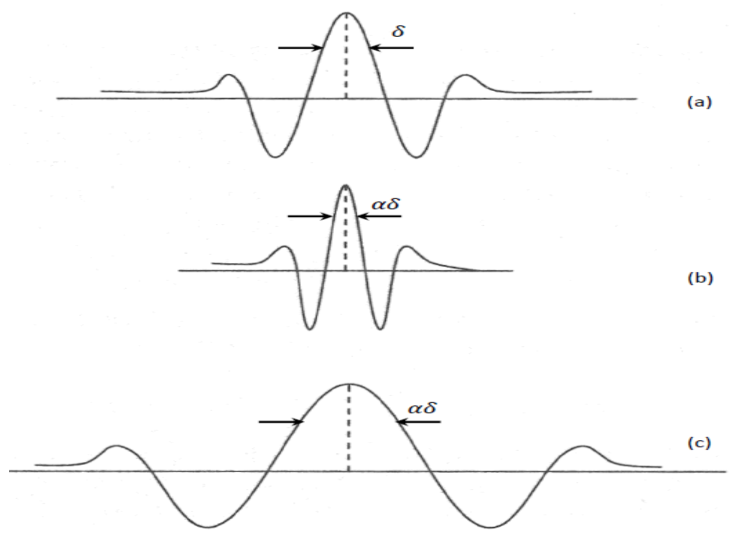


Fig. 12 (a) A mother wavelet, (b) ($/\boldsymbol{\alpha}$) :$\mathbf{0} < \alpha < 1$, (c)$\boldsymbol{\psi}(\ \boldsymbol{t}/\boldsymbol{\alpha})$: $\boldsymbol{\alpha} > 1$[9].

3.4 DISCRETE WAVELET TRANSFORMS

The Wavelet Series is just a sampled version of CWT and its computation may consume significant amount of time and resources, it is depending on the resolution requirement. The Discrete Wavelet Transform (DWT) is based on sub-band coding, it is found to yield a fast computation of Wavelet Transform. Discrete Wavelet Transform is easy to implement and reduces the computation time. The foundations of DWT were in 1976 when it uses to decompose discrete time signals. When similar work was done in speech signal coding which was named as sub-band coding. At 1983, and another technique similar to sub-band coding was developed it is known as pyramidal coding, and later many improvements were made to these coding schemes which is used in efficient multi-resolution analysis schemes of image. In Continues Wavelet Transform, the signals are analyzed using a set of basic functions which relate to each other by simple scaling and translation or shifting. In the case of DWT, time-scale representation of the digital signal is obtained using digital filtering method. The signal to be analyzed is passed through filters with different cutoff frequencies at different scales.

Since the input signal (digital image) is processed by a digital computing machine, it is define the discrete version of the wavelet transform is known as DWT. To define the wavelet in terms of discrete values of the scaling (dilation) and shifting (translation) parameters *a* and *b* instead of being continuous, *a* and *b* make discrete value using the Eq. 3.6,

$$\boldsymbol{a} = \mathbf{a_0^m} \boldsymbol{b} = \boldsymbol{n b_0 a_0^m}$$

Where *m* and *n* are the arbitrary integers. The value *a* and *b* substituted in Eq. 3.1 by Eq. 3.6, the discrete wavelets can be represented by Eq. 3.7.

$$\boldsymbol{\Psi_{m,n}(t) = a_0^{-m/2} \Psi(a_0^{-m} t - n b_0)} \tag{3.7}$$

There are many choices to select the values of $a0$ and $b0$. By selecting $a0$=2 and $b0$=1, a=2^m and b= $n2^m$. This corresponds to sampling (discretization) of *a* and*b* in such a way that the consecutive discrete values of *a* and *b* as well as the sampling intervals differ by a factor of 3. This way of sampling is popularly known as dyadic decomposition, and using these values, it is possible to represent the discrete wavelets as in eq. 3.8, it is constitutes a family of orthonormal basis functions

$$\boldsymbol{\Psi_{m,n}(t) = 2^{-m/2} \Psi(a_0^{-m} t - n b_0)} \tag{3.8}$$

In general, the wavelet coefficients for function $f(t)$ are given by

$$\boldsymbol{C_{m,n}(f) = a_0^{-m/2} \int f(t) \Psi(a_0^{-m} t - n b_0) dt} \tag{3.9}$$

And hence for dyadic decomposition, the wavelet coefficients can be derived accordingly as

$$\boldsymbol{C_{m,n}(f) = 2^{-m/2} \int f(t) \Psi(a_0^{-m} t - n b_0) dt} \tag{3.10}$$

This allows us to reconstruct the signal $f(t)$ in form the discrete wavelet coefficients as

$$\boldsymbol{f(t) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} C_{m,n}(f) \Psi_{m,n}(t)} \tag{3.11}$$

The transform shown in Eq. 3.9 is called the wavelet series, it is analogous to the Fourier series because the input function *f*(t) is still a continuous function whereas the transform coefficients are discrete in nature. This is often called the discrete time wavelet transform (DTWT).

The advantage of the DWT over Fourier transformation is that it performs multi-resolution analysis of signals with localization both in time and frequency domain popularly known as time-frequency localization, and as a result, DWT decomposes a digital signal into different sub-bands so that the lower frequency sub-bands have good frequency resolution and coarser time resolution as compared to the higher frequency sub-bands. Discrete Wavelet Transform is highly used in image compression due to the fact that the DWT supports features like progressive image transmission of by quality and resolution, and ease of image compression coding and manipulation. Because of these characteristics, DWT is the basis of the new JPEG2000 image compression standard.

3.4.1 ONE DIMENSIONAL DWT

Any signal is first applied to a pair of low-pass and high-pass filters. Then down sampling (neglecting the alternate coefficients) is applied to these filtered coefficients. The filter pair (h, g) which is used for decomposition is called analysis filter-bank and the filter pair which is used for reconstruction of the signal is called synthesis filter bank. The output of the low pass filter after down sampling contains low frequency components of the signal which is approximate part of the original signal and the output of the high pass filter after down sampling contains the high frequency components which are called details (i.e., highly textured parts like edges) of the original signal.

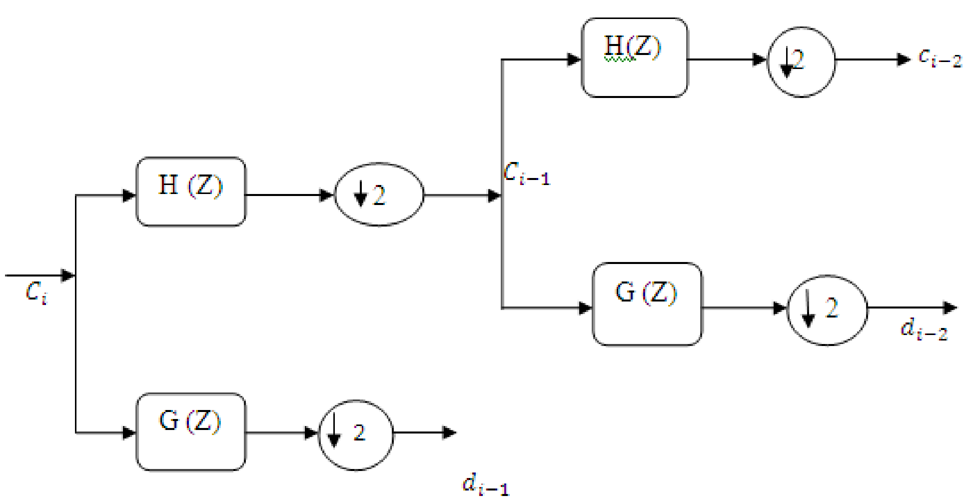


Fig. 13 One dimensional two level wavelet decomposition

This approximate part can be further decomposed into low frequency and high frequency part. This process can be continued successively to the required number of levels of the process. This process is called multi level decomposition, as shown in Fig. 3.4, and in the reconstruction process method, these approximate and detail coefficients are first up-sampled and then applied to low-pass and high-pass reconstruction filters. Then these filtered coefficients are added to get the reconstructed version of the original image or data.

This process can be extended to multi level reconstruction. So that the approximate coefficients of this block have been formed a pairs of approximate and detail coefficients. Fig. 3.3 shown as below

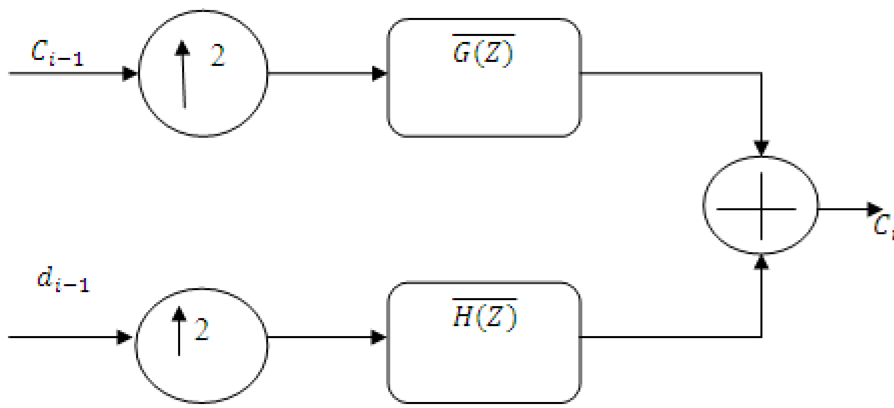


Fig 14 One dimensional inverse wavelet transforms

3.4.2 TWO-DIMENSIONAL DWT

One dimensional DWT can be easily extended to two dimensions which can be used for the transformation of two dimensional images. 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.: 3.7(c). 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, shown in Fig. 3.7. Various architectures have been proposed for computation of the Discrete Wavelet Transform. 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

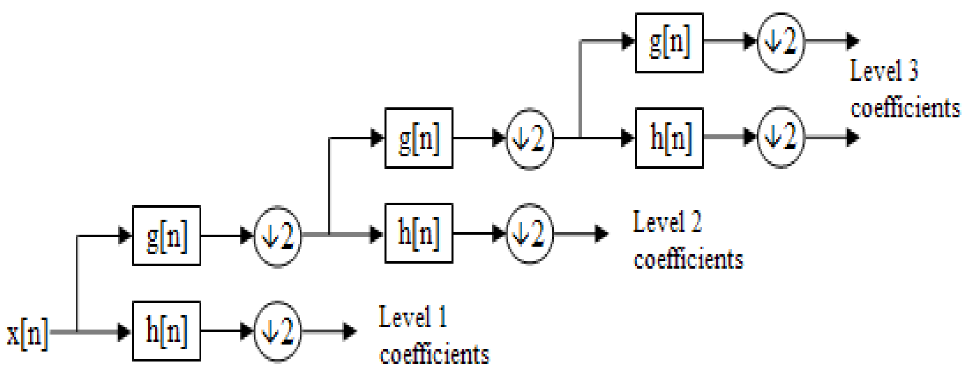
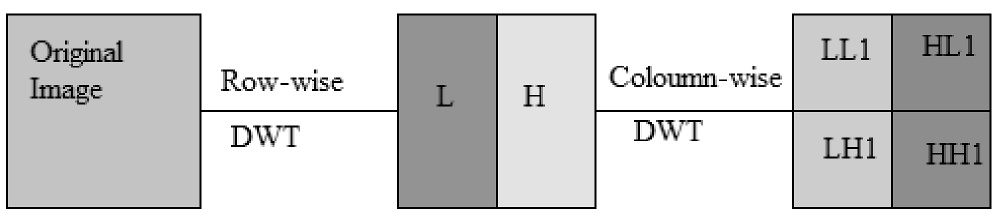
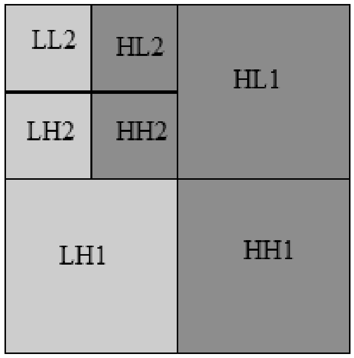


Fig 15 Two channel filter bank at level 3[3]

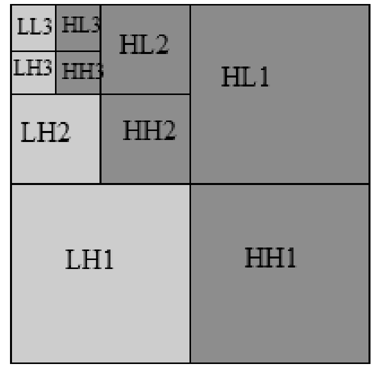


(a) First level Decomposition

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.



(b) Second level Decomposition



(c) Third level Decomposition

Fig. 16 Row-column computation of 2-D DWT

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 tow-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 as shown in fig.3.7

3.5 EXTENSION TO TWO-DIMENSIONAL SIGNALS

The two-dimensional extension of DWT is essential for transformation of two-dimensional signals, such as a digital image. A two-dimensional digital signal can be represented by a two-dimensional array $X[M,N]$ with M rows and N columns, where M and N are nonnegative integers of 2D image array. The simple approach for two-dimensional implementation of the DWT is to perform the one-dimensional DWT row-wise to produce an intermediate result and then perform the same one-dimensional DWT column-wise on this intermediate result to produce the final result. This is shown in Fig. 3.7(a). This is possible because the two-dimensional scaling functions can be expressed as separable functions which is the product of two-dimensional scaling function such as $\Phi_2(x,y) = \Phi_1(x)\Phi_1(y)$. The same is true for the wavelet function $\Psi(x,y)$ as well. Applying the one-dimensional transform in each row of image, two sub-bands are produced in each row. When the low-frequency sub-bands of all the rows (L) are put together, it looks like a thin version (of size $M \times \frac{N}{2}$ of the input signal as shown in Fig. 3.7(a). Similarly put together the high-frequency sub-bands of all the rows to produce the H sub-band of size$M = \frac{N}{2}$, which contains mainly the high-frequency information around discontinuities (edges in an image) in the input signal. So that applying a one-dimensional DWT column-wise on these L and H sub-bands (intermediate result), four sub-bands LL, LH, HL, and HH of size $\frac{M}{2} \times \frac{N}{2}$ are generated as shown in Fig..3.7(a) LL is a coarser version of the original input signal. LH, HL, and HH are the high frequency sub-band containing the detail information of the image. It is also possible to apply one-dimensional DWT column-wise first and then row-wise to achieve the same result. Fig..3.8 comprehends the idea describe above.

The multi-resolution decomposition approach in the two-dimensional signal is demonstrated in Fig.s.3.7 (b) and (c). After the first level of decomposition, it generates four sub-bands LL1, HL1, LH1, and HH1 as shown in Fig. 3.7 (a).

Fig. 17 Extension of DWT in two - dimensional signals

Considering the input signal is an image, the LL1 sub-band can be considered as a 2: 1 sub-sampled (both horizontally and vertically) version of image. The other three sub-bands HL1, LH1, and HH1 contain higher frequency detail information. These spatially oriented (horizontal, vertical or diagonal) sub-bands mostly contain information of local discontinuities in the image and the bulk of the energy in each of these three sub-bands is concentrated in the vicinity of areas corresponding to edge activities in the original image.

3.6 DWT TWO CHANNEL FILTER BANK

The wavelet transform decomposes an image into a set of different resolution sub-image, corresponding to the various frequency bands. Sub-band coding is a procedure in which the input signal is subdivided into several frequency bands. Sub-band coding can be implemented through a filter bank. A filter bank is a collection of filter having either a common input or common output. When the filters have a common input, they form an analysis bank and when they share a common output, they form a synthesis bank. The basic idea in a filter bank is to partition a signal dyadically at the frequency domain. First, let us analysis the perfect reconstruction criteria for a tow channel filter bank for a one dimensional signal, and the same concept can be easily extended to a two-dimensional signal, if the two-dimensional signal is separable.

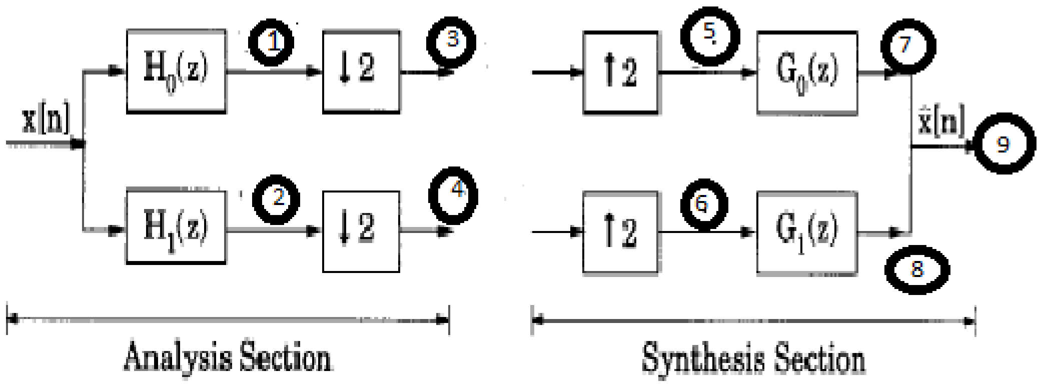


Fig 18 Two-band analysis-synthesis filter bank [3].

The tow-channel filter bank is composed by two section, analysis section and synthesis section, as shown in fig.3.9. The analysis section decomposes the signal into a set of sub-band components and the synthesis section reconstructs the signal from its components. The sub-band analysis and synthesis filters should be designed to be alias-free and are also required to satisfy the perfect signal reconstruction property. The simultaneous cancellation of aliasing as well as amplitude and phase distortion leads to perfect reconstruction filter banks which are suitable for hierarchical sub-band coding and multi-resolution signal decomposition.

The analysis filter bank splits the original image signal two equal frequency bands. Here the filters $\mathrm{H}_0[z]$and $\mathrm{H}_1[z]$act as low-pass and high-pass filters respectively. After filtering, the signal outputs at 1 and 2 are given in Eqn.(3.12) and (3.13) respectively.

At node (1) :- $\mathrm{X}[z]\mathrm{H}_0[z]$ (3.12)

At node (2) :- $\mathrm{X}[z]\mathrm{H}_1[z]$ (3.13)

After filtering, the signal's sampling frequency is too high, and hence half the samples are discarded by the down-sampling operation. After decimation, the Z-transform is given in Eqn. (3.14) and (3.15) respectively.

At node (3):- $\mathrm{Y}[z] = \frac{1}{2}\left\{\mathrm{X}\left[z^{1/2}\right] \cdot \mathrm{H}_0\left[z^{1/2}\right] + \mathrm{X}\left[-z^{1/2}\right] \cdot \mathrm{H}_0\left[-z^{1/2}\right]\right\}$ (3.14)

At node (4):- $\mathrm{Y}[z] = \frac{1}{2}\left\{\mathrm{X}\left[z^{1/2}\right] \cdot \mathrm{H}_1\left[z^{1/2}\right] + \mathrm{X}\left[-z^{1/2}\right] \cdot \mathrm{H}_1\left[-z^{1/2}\right]\right\}$ (3.13)

The synthesis filter bank reconstructs the signal from the two filtered and decimation signals. The synthesis procedure involves expending the signals in each branch by two which is termed expansion or interpolation. The interpolation is achieved by inserting zeros between successive samples. After interpolation, the Z-transform of the signal at node 3 and 6 are given in Eqn. (3.16) and (3.17) respectively.

At node (3):- $\mathrm{X}[z] = \frac{1}{2}\{\mathrm{X}[z] \cdot \mathrm{H}_0[z] + \mathrm{X}[-z] \cdot \mathrm{H}_0[-z]\}$ (3.16)

At node (6):- $\mathrm{X}[z] = \frac{1}{2}\{\mathrm{X}[z] \cdot \mathrm{H}_1[z] + \mathrm{X}[-z] \cdot \mathrm{H}_1[-z]\}$ (3.17)

The above Eqn. (3.16) and (3.17) can be written in matrix form as given below

$$\frac{1}{2} \times \begin{bmatrix} \mathrm{H}_0[z] & \mathrm{H}_0[-z] \\ \mathrm{H}_1[z] & \mathrm{H}_1[-z] \end{bmatrix} \begin{bmatrix} \mathrm{X}[z] \\ \mathrm{X}[-z] \end{bmatrix} \quad (3.18)$$

At node 7 and 8

$$\frac{1}{2} \times [\mathrm{G}_0[z] \quad \mathrm{G}_1[z]] \cdot \begin{bmatrix} \mathrm{H}_0[z] & \mathrm{H}_0[-z] \\ \mathrm{H}_1[z] & \mathrm{H}_1[-z] \end{bmatrix} \cdot \begin{bmatrix} \mathrm{X}[z] \\ \mathrm{X}[-z] \end{bmatrix} \quad (3.19)$$

$$\frac{1}{2} \times [\mathrm{G}_0[z] \quad \mathrm{G}_1[z]]_{1\times 2} \cdot \begin{bmatrix} \mathrm{H}_0[z] & \mathrm{H}_0[-z] \\ \mathrm{H}_1[z] & \mathrm{H}_1[-z] \end{bmatrix}_{2\times 2} \cdot \begin{bmatrix} \mathrm{X}[z] \\ \mathrm{X}[-z] \end{bmatrix}_{2\times 1} \tag{3.20}$$

Combining both G and H matrices using matrix multiplication:

$$\left[\frac{\mathrm{G}_0[z]\mathrm{H}_0[z]+\mathrm{G}_1[z]\mathrm{H}_1[z]}{2} \; \frac{\mathrm{G}_0[z]\mathrm{H}_0[-z]+\mathrm{G}_1[z]\mathrm{H}_1[-z]}{2}\right]_{1\times 2} \cdot \begin{bmatrix} \mathrm{X}[z] \\ \mathrm{X}[-z] \end{bmatrix}_{2\times 1} \tag{3.21}$$

$$\mathrm{F}_0[z] = \frac{\mathrm{G}_0[z]\mathrm{H}_0[z]+\mathrm{G}_1[z]\mathrm{H}_1[z]}{2} \tag{3.22}$$

$$\mathrm{F}_1[z] = \frac{\mathrm{G}_0[z]\mathrm{H}_0[-z]+\mathrm{G}_1[z]\mathrm{H}_1[-z]}{2} \tag{3.23}$$

So that

$$\left[\mathrm{F}_0[z]\mathrm{F}_1[z]\right]_{1\times 2} \cdot \begin{bmatrix} \mathrm{X}[z] \\ \mathrm{X}[-z] \end{bmatrix}_{2\times 1} \tag{3.24}$$

$$\mathrm{F}_0[z]\mathrm{X}[z] + \mathrm{F}_1[z]\mathrm{X}[-z] \tag{3.25}$$

In the above equation,$\mathrm{X}[-z]$ refers to the aliasing component. This aliasing will spoil the signal. So select the filter co-efficient in order to reduce the aliasing effect of sample, that's make the $\mathrm{F}_1[z]$as zero to neglect the aliasing effect. Let,

$$\mathrm{H}_0[z] = H[z]; \qquad \mathrm{H}_1[z] = H[-z]; \qquad \mathrm{G}_0[z] = 2H[z]; \qquad \mathrm{G}_1[z] = 2H[-z]; \tag{3.26}$$

From the above conclusion, we can say that the four filter designs are given by a single filter co-efficient. This is the beauty of sub-band coding.

When we substitute the above assumptions:

$$\mathrm{F}_1[z] = \frac{\mathrm{G}_0[z]\mathrm{H}_0[-z] + \mathrm{G}_1[z]\mathrm{H}_1[-z]}{2} => \; 0$$

$$\mathrm{F}_0[z] = \frac{\mathrm{G}_0[z]\mathrm{H}_0[z] + \mathrm{G}_1[z]\mathrm{H}_1[z]}{2}$$

$$= \frac{2H[z]H[z] + (-2H[-z])H[z]}{2}$$

$$= H^2[z] + H^2[-z]$$

So, finally at node 9

$$(H^2[z] + H^2[-z]) \cdot \mathrm{X}[z] \tag{3.27}$$

While transmitting from one place to another, the delay is unavoidable though the delay value may be mille-seconds. For a perfect reconstruction filter bank, the reconstruction signal is the delayed version of the original signal which is given by,

$$(H^2[z] + H^2[-z]) \cdot \mathrm{X}[z] = z^{-k} \cdot \mathrm{X}[z] \tag{3.28}$$

$$(H^2[z] + H^2[-z]) = z^{-k}$$

That is, $\mathrm{H}[z] = \mathrm{A}[z] \cdot z^{-\left(\frac{N-1}{2}\right)}$

Then the signal value at node (9) is given by

$$A^2[z] \cdot z^{-(N-1)} - A^2[-z] \cdot (-z)^{-(N-1)} = z^{-k}$$

$$A^2[z] \cdot z^{-(N-1)} - A^2[-z] \cdot (-1)^{-(N-1)} \cdot (z)^{-(N-1)} = z^{-k}$$

$$A^2[z] \cdot z^{-(N-1)} - A^2[-z] \cdot (z)^{-(N-1)} \cdot (-1)^{-(N-1)} = z^{-k}$$

If k=N-1 (delay is governed by the filter co-efficient)

$$A^2[z] \cdot z^{-(N-1)} - A^2[-z] \cdot (z)^{-(N-1)} \cdot (-1)^{-(N-1)} = z^{-(N-1)}$$

$$A^2[z] - \left(A^2[-z] \cdot (-1)^{-(N-1)}\right) = 1$$

If N is even (for PR condition)

$$A^2[z] + A^2[-z] = 1$$

$$H^2[z] + H^2[-z] = 1$$

Then the condition for perfect reconstruction is given by

$$H^2[z] + H^2[-z] = 1 \tag{3.29}$$

3.7 DESIRABLE CHARACTERISTICS OF A FILTER BANK

The desirable characteristics of a filter bank include maximal decimation, separable filtering, polyphase form, perfect reconstruction, and tree structure

(i) Maximal Decimation

The maximal decimation property indicates that the number of coefficients produced by decomposition is the same as the number of input samples. This is also known as critical sampling. This property is associated with computational complexity as it keeps the number of samples to be processed to a minimum.

(ii) Separable Filtering

Separable filtering indicates that two-dimensional as well as higher dimensional filtering can be performed as one-dimensional filtering. For example, two-dimensional filtering is performed as two one-dimensional filtering performed row-wise and column-wise. This property is associated with computational efficiency as separable filtering is more efficient than the equivalent non-separable filtering

(iii) Polyphase Form

Polyphase form is efficient implementation of decimated filter bank where the filtering is performed after the down-sampling, thus reducing the number of components.

(iv) Perfect Reconstruction

Perfect reconstruction property ensures that the reconstructed signal resembles the original signal without any error. That is, the coefficient produced by the forward transform can be sent through the inverse transform to reproduce the original signal without any error.

(v) Tree-Structure Decompositions

Tree structure decomposition sub-divides radically the low frequency region. Such decomposition are well-suited to processing of natural images which tend to have the energy concentrated in radically low-frequency regions.

3.8 DWT ALGORITHM FOR TEXT EXTRACTION

Wavelet analysis can be used divided the information of an image into approximation and detailed sub image signal. The approximation sub signal shows the generally pixel value of image, and three detailed sub signal show horizontal, vertical and diagonal details (changes in image). Otherwise if these detail is very small than they can be set to zero without significantly changing the picture.

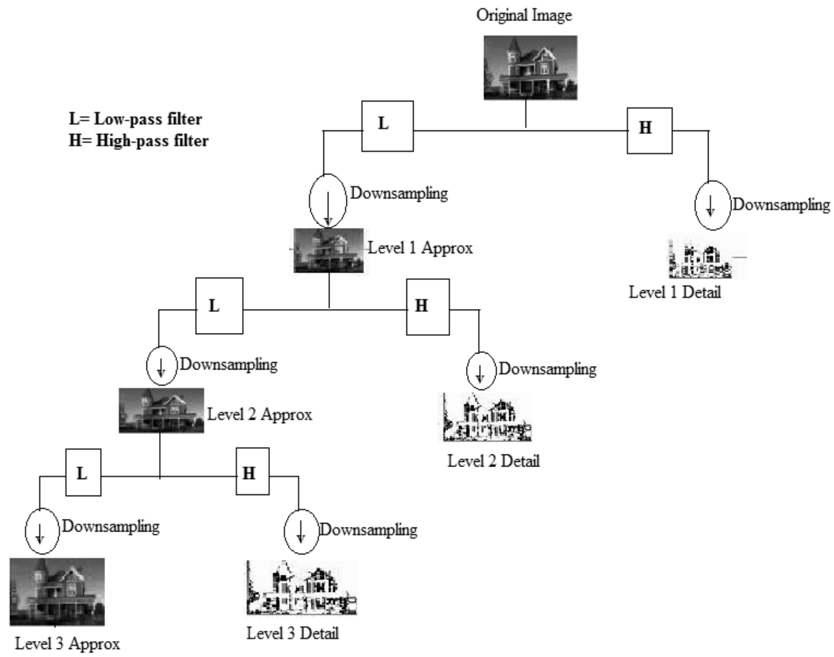


Fig 19: Decomposition of an input image using a wavelet transformation using three passes

If the number of zeroes is greater than the compression ratio is also high. There is two types of wavelet is used in image compression. First one is Continues wavelet transform and second one is discrete wavelet transform. The Wavelet analysis is computed by filter bank. This is combination of high-pass and low-pass filters. High pass filter kept high frequency information and lost low frequency information. Low pass filter kept law frequency information and lost high frequency information.

So signal is effectively decomposed into two parts, a detailed part (high frequency) and approximation part (law frequency). The Level 1 detail is horizontal detail, the level2 detail is vertical detail and level 3 details is diagonal detail of the image signal. The Flow chart representation of DWT algorithm for image compression using VHDL shown in Fig.3.12, according HAAR DWT algorithm, first applying reset signal is one then run the simulator, so all the value of the previous input and output will be zero. After then applying a clock pulse on the clock signal and the reset signal will be zero, all above condition will be done after then the original 2D image will be convert the set of pixels

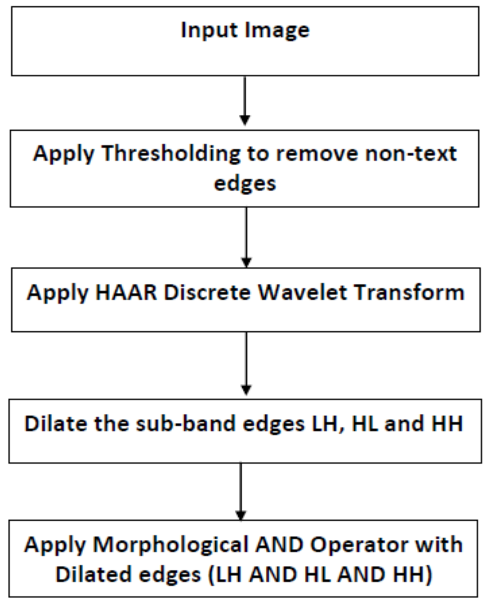


Fig 20 Flow chart representation of DWT algorithm for Text extraction

Every pixels of the 2D image have own x-axis and y-axis, so we will represent the image pixels in histogram representation. After then the image will be applying to a filter bank, the filter bank will consist of Low-pass and High-pass filters, then the image signal will be separated high band signal and low band signal, according the HAAR DWT algorithm the low band and high band image signal have four possible combination, such as LL,LH,HL,HH. The LL band is more significant band it contains more information of the original image, so it is most important part of the algorithm process. The LL band again sub divided to lower band till to the desired output will not obtained, this process shown below in the Fig.3.12

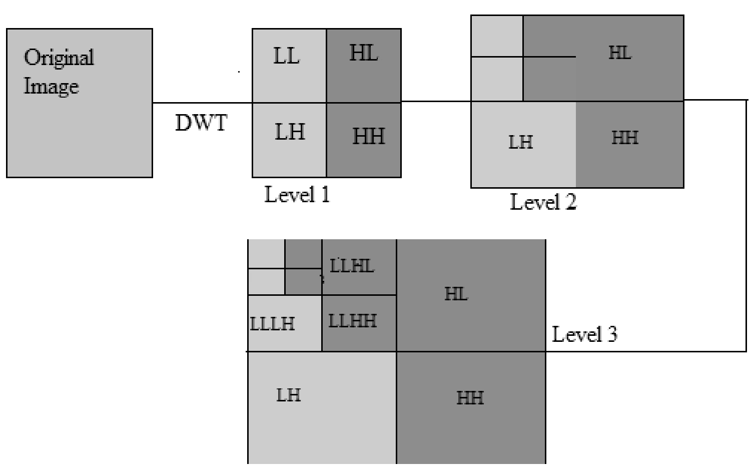


Fig 21 Decomposition of wavelet transforms

3.8.1 ENCODING PROCESS

In discrete wavelet transform six major steps of image compression. These encoding processes are following step-by-step.

Step1

first the original image passed through a combination of filter, such as low-pass and high-pass filter. These filters are applying each row.

Step2

Then output image of the Bothe low-pass and high-pass filter is L1 and H1, these are combining into$\mathrm{T}_1 = [\mathrm{~L1~H1~}]$.

Step3

After the filtering the combine output T_1 of these filters are down sampled by the 3.

Step4

Now, again T1 has been passed through high pass filter and low filter by applying on each column.

Step5

Let suppose the output of the step4 is L2 and H3. Then H2 and L2 combine into $\mathrm{T}_3 = \begin{bmatrix} \mathrm{L2} \\ \mathrm{H2} \end{bmatrix}$

Step6

After the filtering the combine output T_3 of these filters are down sampled by the 3. This is compressed image of the processing.

In Fig. 3.12 there are shown a resulted image after applying encoding process. In this Fig. are four blocks, the first half upper block show the approximation, and second upper half block show the horizontal detail. First lower level block shows vertical detail and second lower level block shows diagonal detail. In this algorithm is shown one level discrete wavelet transform. By applying this process more than one time it can increase the level of DWT. Second and third level DWT gives the better compression ratio of image. But it will come with loss of some information.

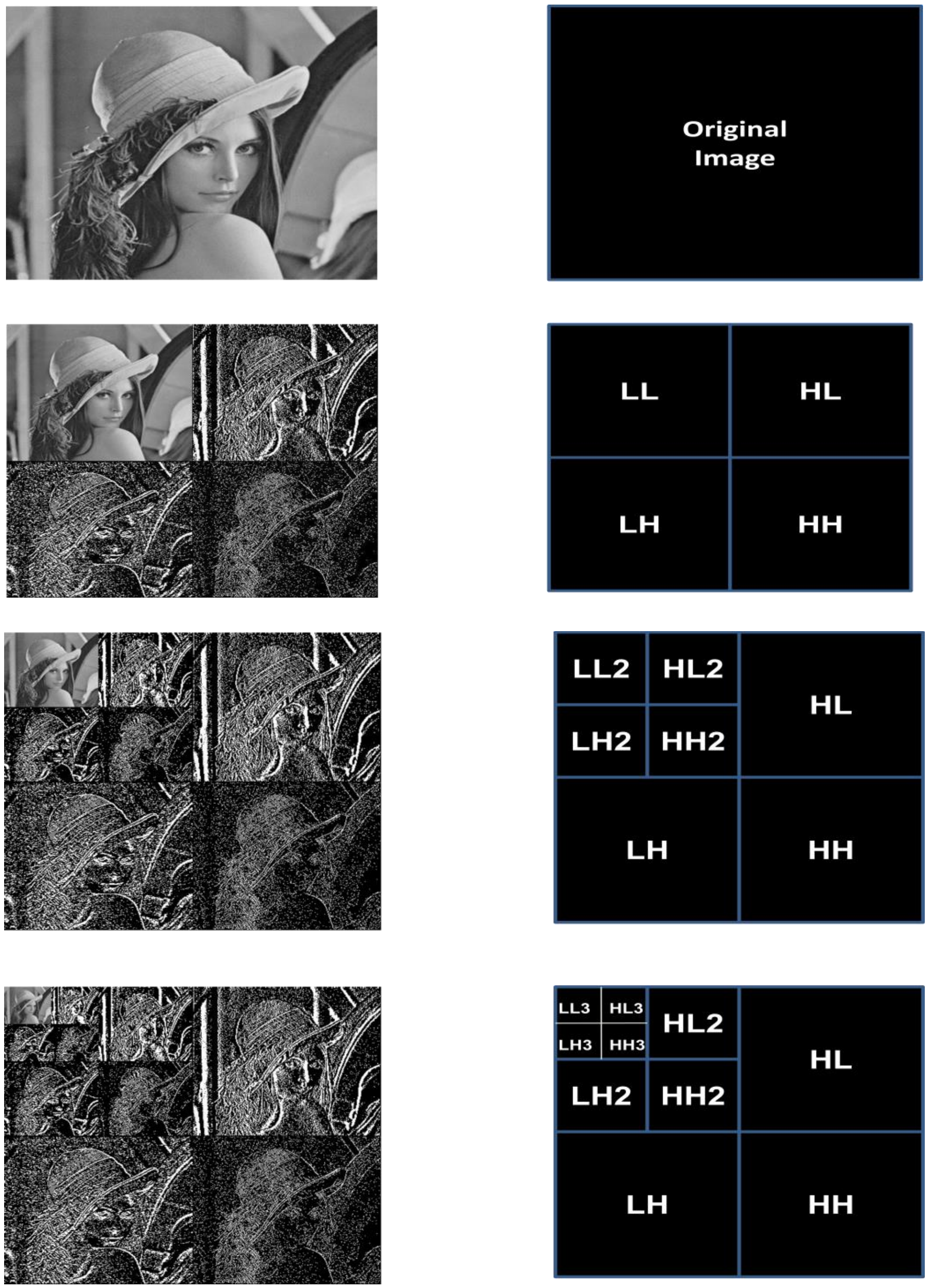


Fig. 22 Texture Compressed Image

3.8.2 DECODING PROCESS

The decoding process is not exact reverse of the encoding process. The steps of the decoding process are following.

Step1

First extract low pass filter image and high pass filter image from the compressed image. The low pass filter image is taking by upper half rectangle matrix and high pass filter image is taking by down half rectangle matrix.

Step 2

These images are up sampled by 3.

Step 3

Then the summation of both images take into one image is called R_1.

Step 4

Then again extract low pass filter image and high pass filter image by simply dividing vertically part of the image. First half is low pass filtered image part. And second half is high pass filter image.

Step 5

After then above process take summation of both images. It is the output of reconstructed image.

So in the DWT get very high extraction ratio, and also lose minimum amount of information. But if more than one level then get more extraction ratio but the reconstructed image is not identical to original image.

3.8.3 TEXTURE ANALYSIS

The discrete wavelet transform is a very useful tool for signal processing and image analysis especially in multi-resolution representation. In DWT signals are decomposed into different components in the frequency domain. 1-D DWT decomposes an input sequence into two components the average component and the detail component by calculations with a low-pass filter and a high-pass filter [9]. Two-dimensional discrete wavelet transform (2-D DWT) decomposes an input image into four sub-bands, one average component (LL) and three detail components (LH, HL, HH) as shown in figure 9. In image processing, the multi-resolution of 2-D DWT has been employed to detect edges of an original image. However, 2-D DWT can detect three kinds of edges at a time while traditional edge detection filters cannot. As shown in figure 10the 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.

LL	HL
LH	HH

Fig. 23 2-D DWT decomposition scheme

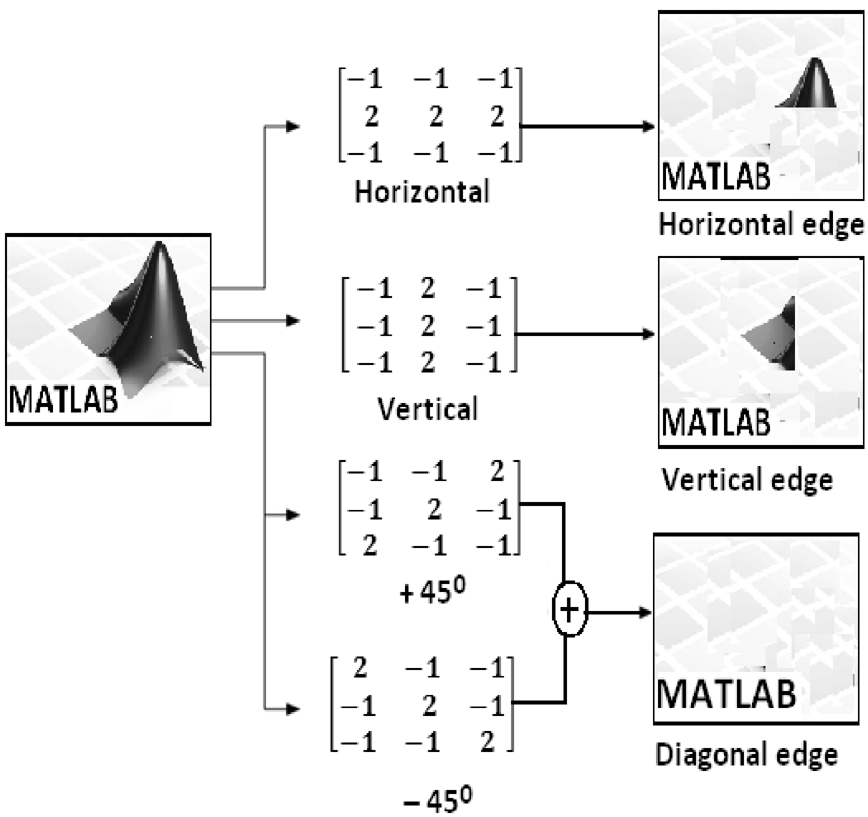


Fig 24 Edge detection using mask operation

Figure 13 shows a gray level image. The 9-7 taps DWT filters decompose this gray image into four sub-bands as shown in Figure 13. 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. Haar DWT has the following important features [17].

- Haar wavelets are real, symmetric and orthogonal.
- Its boundary conditions are the simplest among all wavelet-based methods.
- The minimum support property allows arbitrary spatial grid intervals.
- It can be used to analyze texture and detect edges of characters.
- The high-pass filter and the low-pass filter coefficients are simple (either 1 or –1).

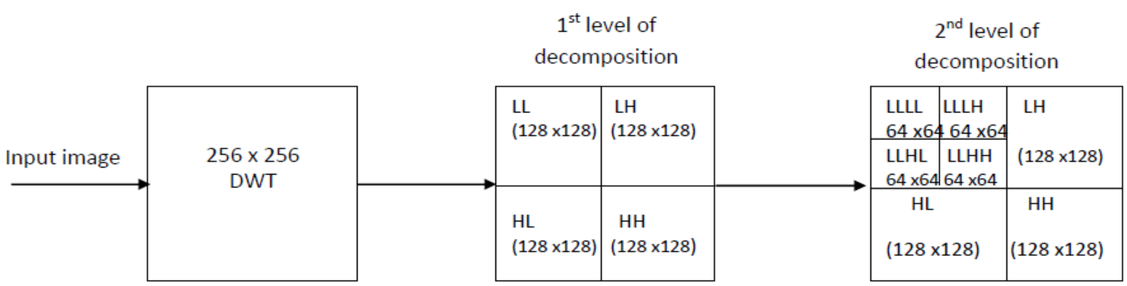


Fig 25 Decomposition scheme of Haar Transform

Figure 16 shows an example of 236 x 236 images, in which it is divided into 128 x 128 LL, LH, HL and LL bands. Further the image is divided into 64 x 64 LLLL, LLLH, LLHL and LLHH bands as shown in figure 17(a) and (b). Figure 18 (a) matrix shows the example of a 4×4 colour image.

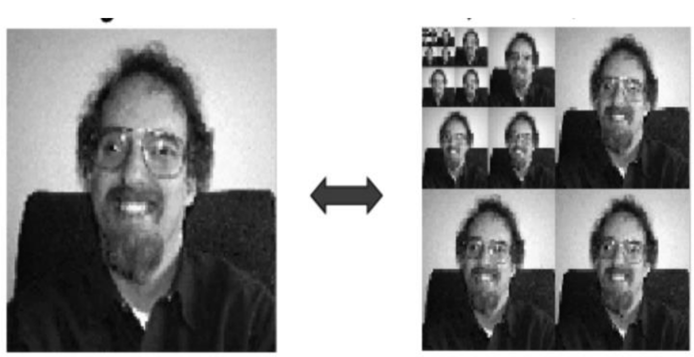


Fig 26 (a) Original image (b) DWT coefficients

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

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

Fig 27 Matrixes (a) The original image (b) the row operation of 2-D Haar DWT and (c) column

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 result shown in figure 18 (b) matrix. Then it is transformed by the column operation and the final resulted 2-D Haar DWT is shown in Figure 18(c) matrix. 2-D Haar DWT decomposes a gray-level image into one average component sub-band and three detail component sub-bands.

- Wavelet function can be freely chosen, no need to divide the input coding into non-overlapping 2-D blocks, it has higher compression ratios avoid blocking artifacts.
- Allows good localization both in time and spatial frequency domain
- Better identification of which data is relevant to human perception because higher compression ratio

CHAPTER 4
RESULTS & DISCUSSIONS

4.1 SIMULATION

In the modelsim waveforms, image_in_x_axis and image_in_y_axis, represents the integer value of image at discrete points at X, and Y axis respectively which is a matrix of 64 pixels values. Sample_x_axis and Sample_y_axis, represents the integer value of image in discrete samples at X and Y axis respectively by which image should be modeled. P_state and n_state are the present state and next state to develop the chip using Finite State Machine (FSM). The RTL view of the developed and synthesized chip is shown in fig. 11 which has the details of the pins of the designed chip. Table 1 list the functionality of the chip.

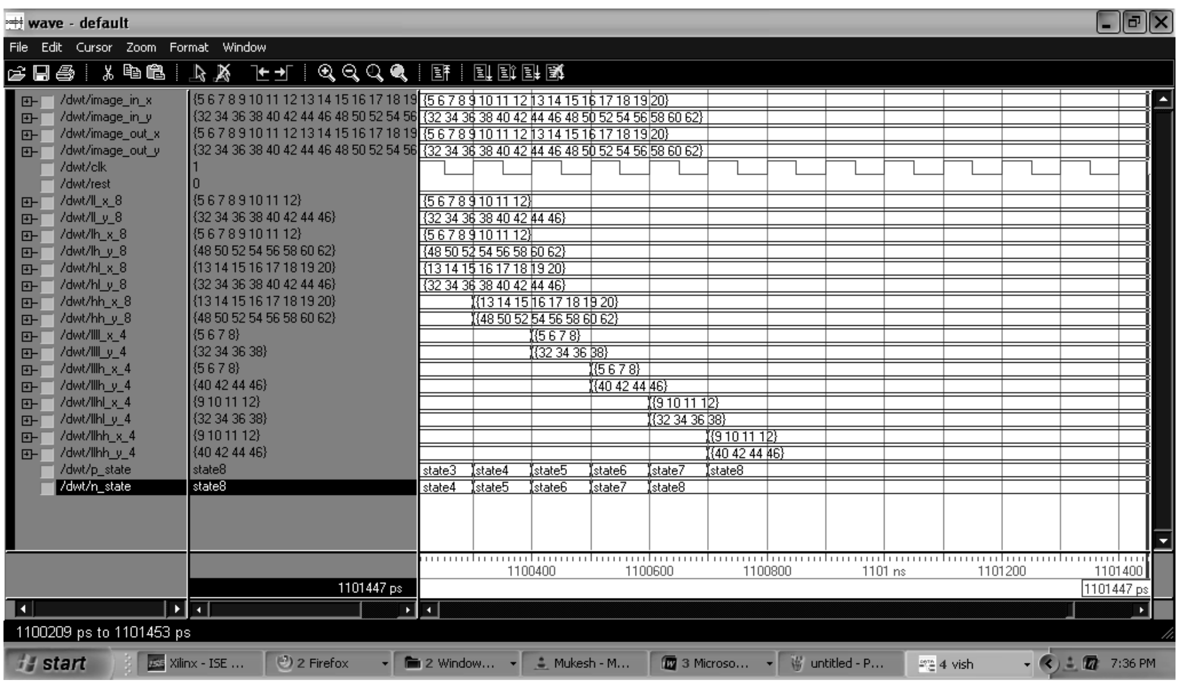


Fig 29: ModelSim Simulation of 2D Image

Fig 30 RTL view of 2D-DWT

Table 1: Pin details of DWT chip for (64 x 64)

Pins	**Functional Description**
reset	Used to reset the memory contents zero for synchronization of the components by using clk of std_logic (1 bit)
clk	Default input for sequential logic to work on rising edge of clock pulse of std_logic.(1 bit)
Image_in_x	Input array to the image in the form of pixels intensity value matrix (integer type in x axis)
Image_in_y	Input array to the image in the form of pixels intensity value matrix (integer type in y axis)
Image_out	Output array to the image in the form of pixels intensity value matrix (integer type decomposed in LL, LH, HH and HL subbands)

4.2 SYNTHESIS RESULTS

Synthesis results are represented by the device utilization reports and timing summary reports. Device utilization report gives the percentage utilization of device hardware for the chip implementation. Device hardware includes No of slices, No of flip flops, No of input LUTs, No. of bounded IoBs, and No of gated clocks (GCLKs) used in the implementation of design. Timing summary and details provides the information of delay, minimum period, maximum frequency, and minimum input arrival time before clock and maximum output required time after clock. Table 2 and table 3 list the synthesis results as device utilization and timing parameters for 2D DWT chip for text extraction. Total memory utilization required to complete the design is also listed in the table. The target device is: xc4vlx20t-2-ff323 synthesized with Virtex-6 FPGA. The synthesis process is carried on Virtex-6 FPGA, in which two 9-pin RS-232 ports assist in the transmission of serial data to and from the FPGA board. In Virtex -4, there is a 40 MHz clock oscillator is the system clock provides the clock signal to the various events taking place within the FPGA and the various programs that require clock for their working. The input switches are given the image_in_x [4:0] and image_in_y [4:0], reset and clock input to the FPGA board, and the results are logged on host computer via an 8-lane PCI Express connection supporting sustained bandwidths of up to 2.38GBytes/sec. The synthesis is carried out simultaneously on 33 images and execution time is noticed. Inbuilt ADC converts the pixels intensity in discrete value, accepted by FPGA and then to integer values with the help of inbuilt DAC. The detailed synthesis process is beyond the scope of the research paper.

Table 2: Device utilization for 2D DWT (64 x 64)

Device Part	**Utilization**	
Number of Slices	11 out of 4656,	2%
Number of Slice Flip Flops	18 out of 9312,	3%
Number of 4 input LUTs	196 out of 493,	20%
Number of bonded IOBs	50 out of 232,	40%
Number of GCLKs	2 out of 24,	3%

Table 3: Timing parameters for 2D DWT (64 x 64)

Timing Parameter	**Utilization**
Minimum period	No path found
Maximum Frequency	400 MHz
Minimum input arrival time before clock	1.731ns
Maximum output required time after clock	4.114ns
Total Memory usage	268688 kB

4.3 FPGA RESULTS

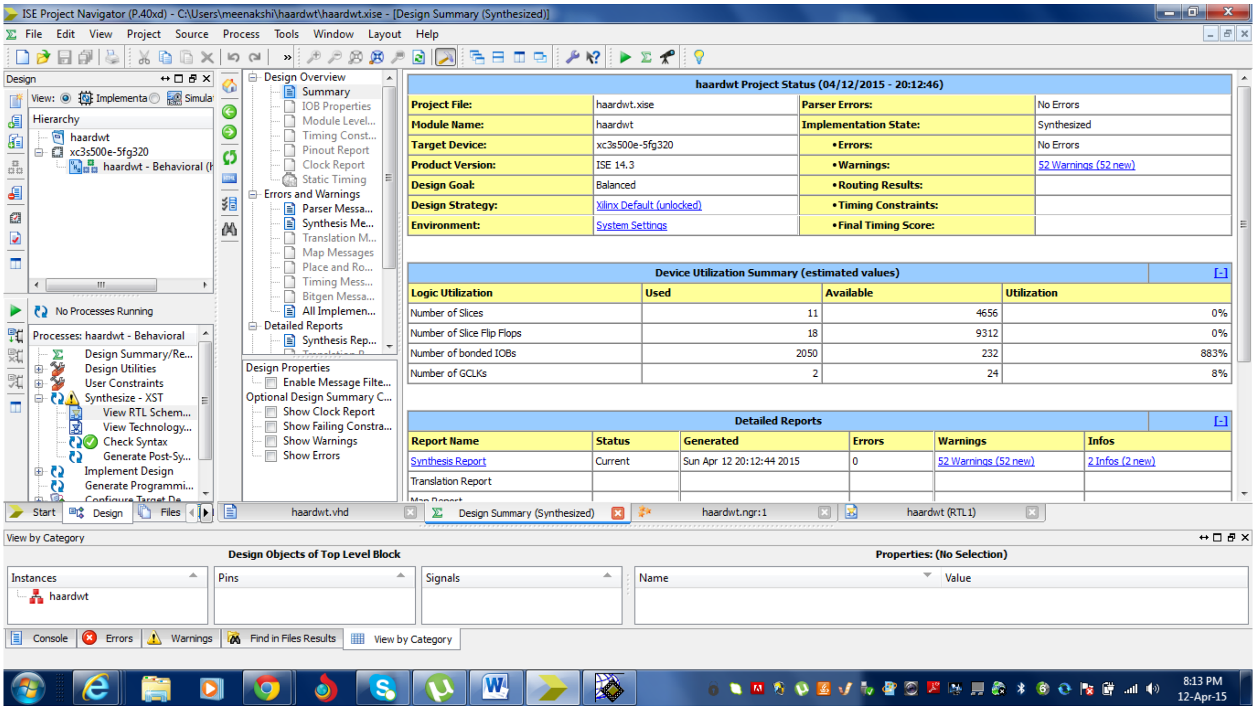


Fig 31 FPGA Synthesis

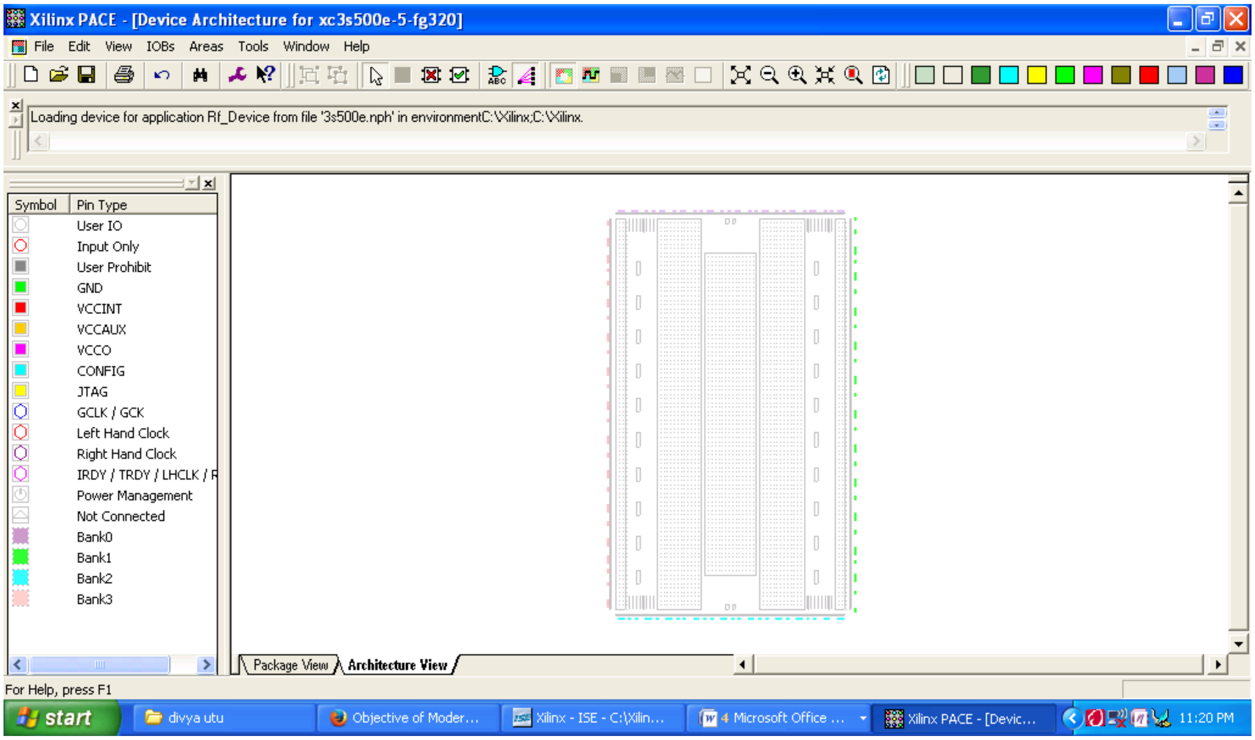


Fig 32 Architecture View

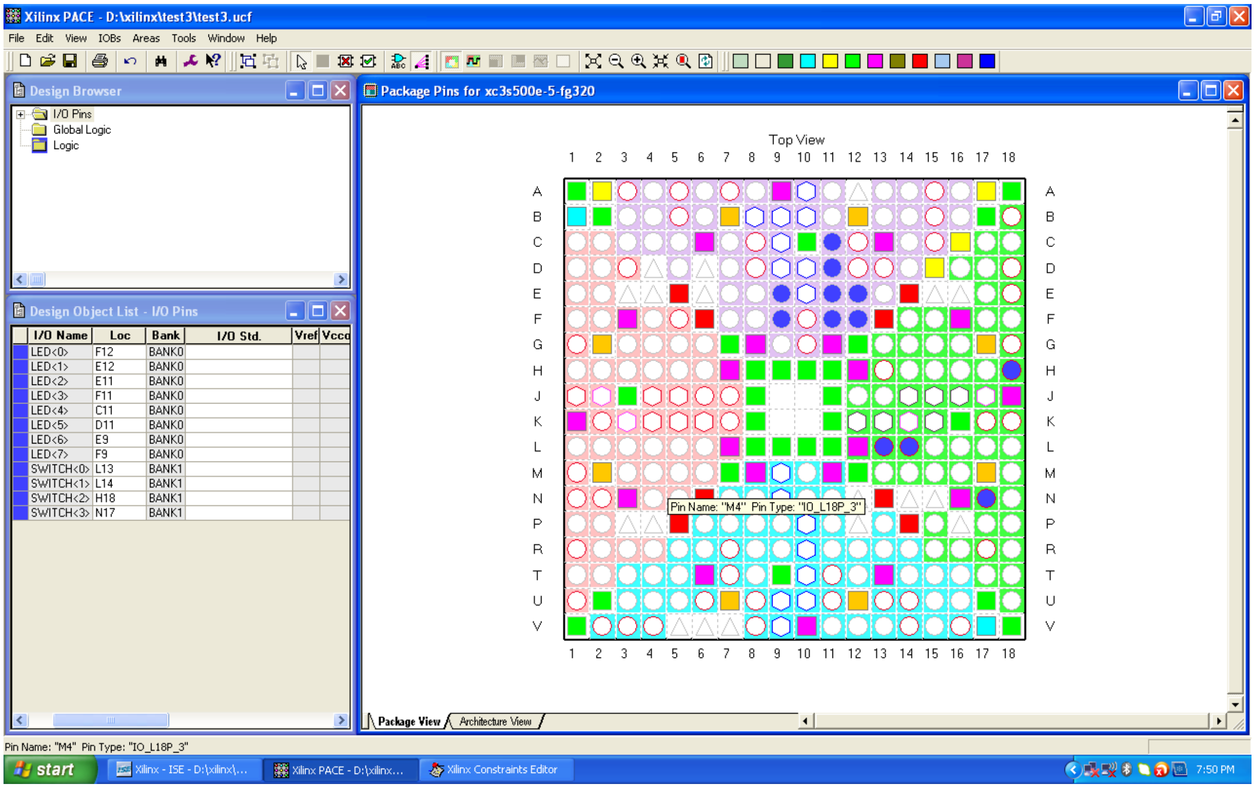


Fig 33 Package View

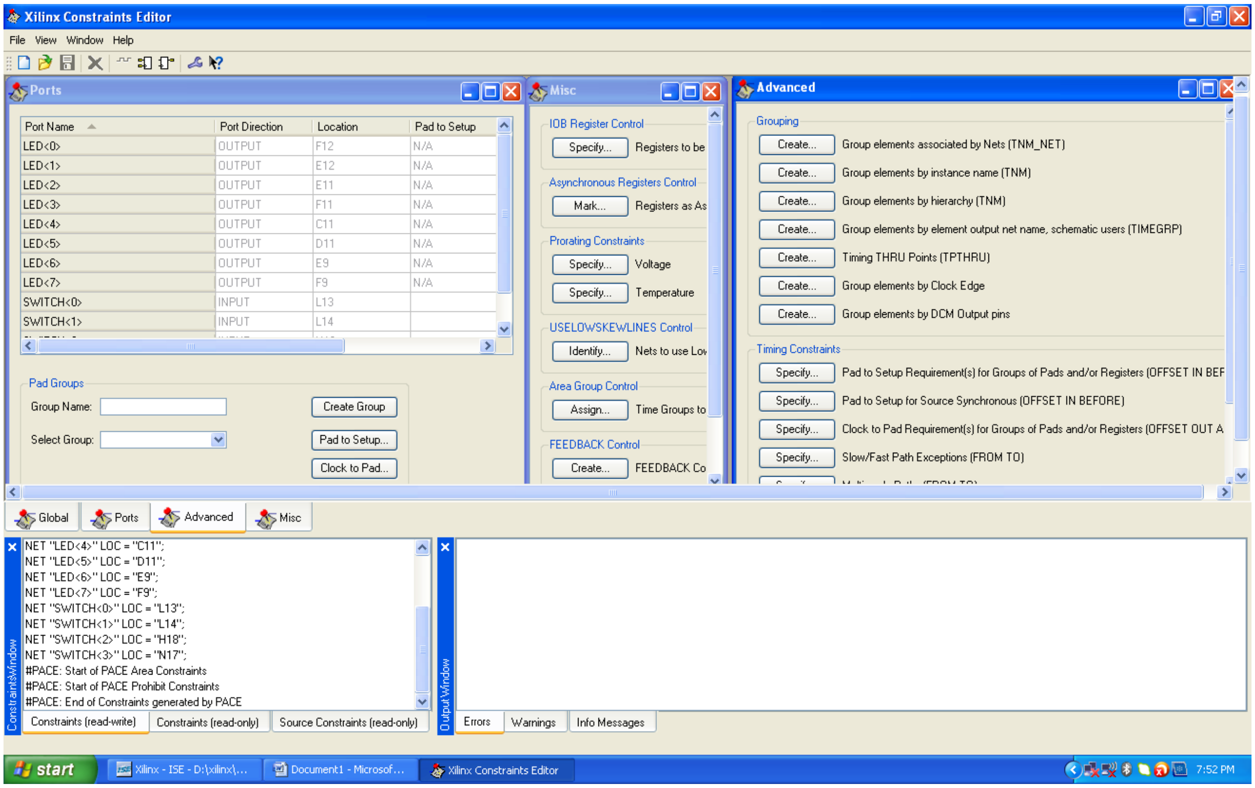


Fig 34 Assignment of Ports

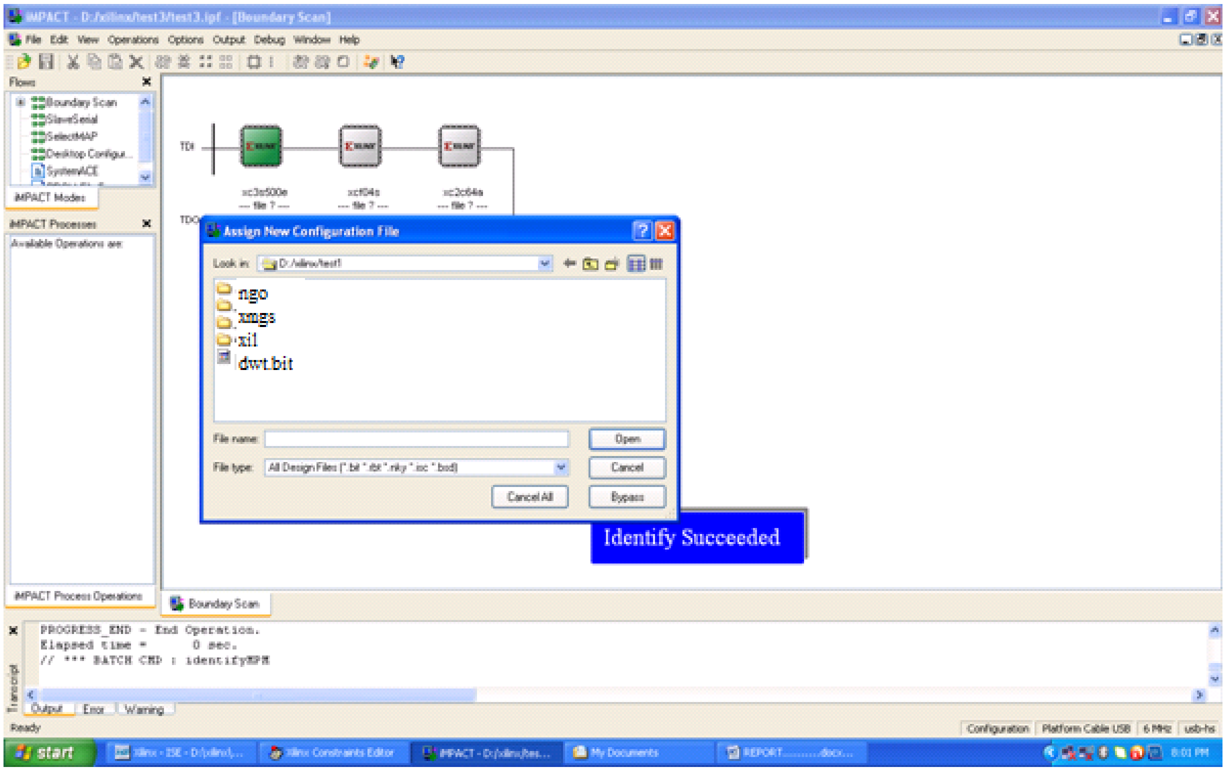


Fig 35Assignment Of Configuration File & Cable Detection

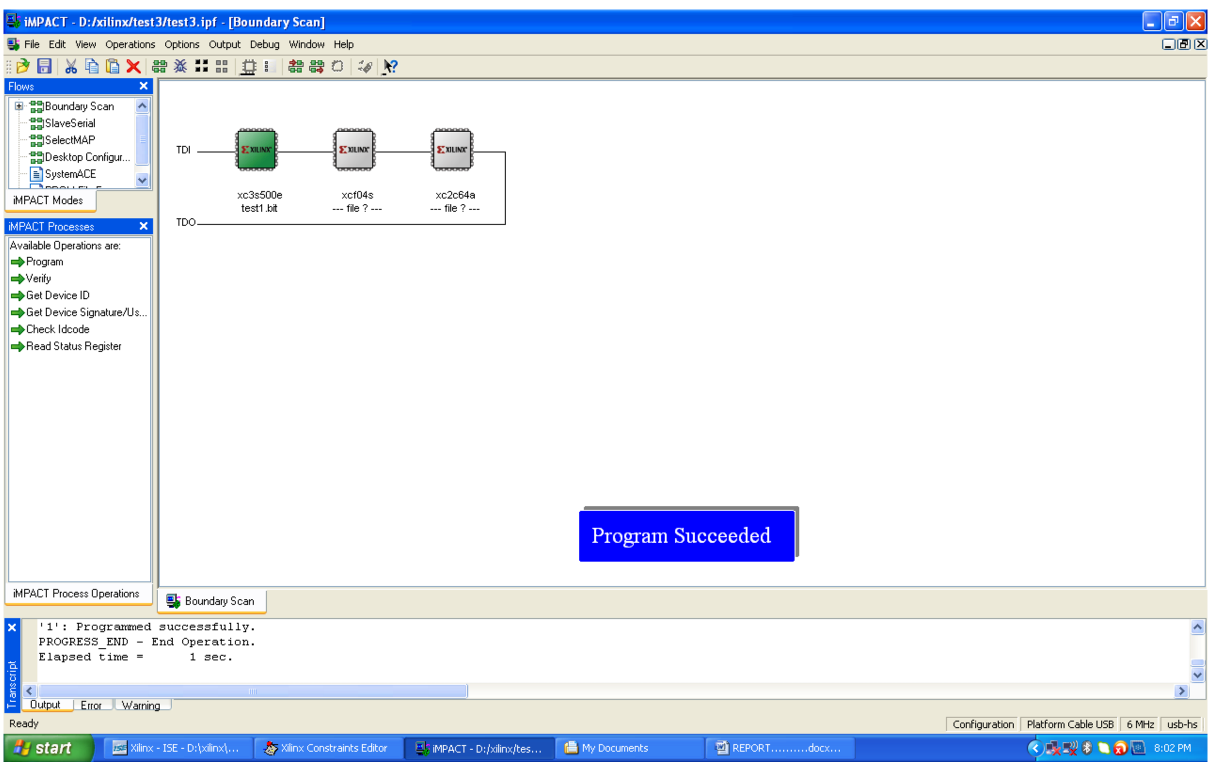


Fig 36 Final Result

CHAPTER 5
CONCLUSION

HAAR DWT has been proved a novel process of text extraction considering multiple cases of image with its textual contents. The algorithm is based on segmenting the image into different regions to extract the textual information may be in horizontal, vertical, inclined. It also uses the methodology of sliding window for reading sub bands of high frequency. The simulation work is carried out for text regions are refined with the integration of mathematical operations such as dilation, erosion, closing and opening in the developed chip. VHDL code is used to develop the chip and synthesis on Virtex-5 FPGA is carried out successfully and results are shown on host computer.

The methodology is very much successful to extract the textual information from a documentary image and results are also analyzed with the help of MATLAB. The work is compared with IEEE ref papers and it is found that the text extraction using HAAR DWT is taken in less time in comparison to the methods discussed in ref paper. The 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 form a video scene. In the addition of the future work can be the integration of DWT chip with the extraction of special characters and integration of the concept of cryptographic techniques of encryption and decryption with larger size of block text and key text.

The text extraction on the colour images using mathematical morphology and Haar DWT is done successfully. Applications of text extraction are huge including the making of digital copies of the ancient scripture to everyday life bills etc. It may be required to be of digital form. 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 form a video scene. Major application of it is that it can be used in license plate recognition of a vehicle as in an approach to determine license plate number.

In the future work we can explore the techniques to recognize the special characters from colour images.In the addition of the future work can be the integration of DWT chip with the extraction of special characters and integration of the concept of cryptographic techniques of encryption and decryption with larger size of block text and key text.

REFERENCES

[1] Xiaoqing Liu and JagathSamarabandu, An Edge-based text region extraction algorithm for Indoor mobile robot navigation, Proceedings of the IEEE, July 2005.

[2] Xiaoqing Liu and Jagath Samarabandu, Multiscale edge-based Text extraction from Complex images, IEEE, 2006.

[3] JulindaGllavata, Ralph Ewerth and Bernd Freisleben, A Robust algorithm for Text detection in images, Proceedings of the 3rd international symposium on Image and Signal Processing and Analysis, 2003.

[4] 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.

[5] Kongqiao Wang and Jari A. Kangas, Character location in scene images from digital camera, The journal of the Pattern Recognition society, March 2003.

[6] 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 17thInternational Conference on Pattern Recognition (ICPR '04), IEEE.

[7] Victor Wu, RaghavanManmatha, and Edward M. Riseman, TextFinder: An Automatic System to Detect and Recognize Text in Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 21, No. 11, November 1999.

[8] 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.

[9] 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.

[10] 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.

[11] R.C. Gonzales and R.E. Woods, Digital Image Processing, Addison-Wesley, Reading, 1992.

[12] J.D. Foley, A. van Dam, S.K. Feiner and J.F. Hughes, Computer Graphics, Principles and Practice, Addison-Wesley, Reading, 1990.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] W. Sowmya, Roy PailyIn, "FPGA implementation of Image Enhancement Algorithm" International conference on Communication and signal processing(ICCSP) IEEE pp 584-588, 2011

[18] K. Sri Rama Krishna, A. Guruva Reddy, M. N. Giri Prasad "A Fault Tolerant FPGA based image Enhancement Filter using Self Healing Algorithm" International Journal of Engineering Science and Technology Vol 2, pp 4922-4928 2010.

[19] ArnabSinha and SumanaGupta,"A Fast Nonparametric Non causal MRF-Based Texture Synthesis Scheme Using a Novel FKDE Algorithm" IEEE Transactions on Image Processing, No.3, March 2010.

[20] Yuichi Tanaka,Madoka Hasegawa, Shigeo Kato, Masaaki Ikehara, and Truong Q. Nguyen, "Adaptive Directional Wavelet Transform Based on Directional Pre filtering", IEEE Transactions on Image Processing, Volume 19, No.4, April 2010.

ANNEXURE

Adesh Kumar, Prakshi Rastogi**, Pragyan Srivastava** " Design and FPGA implementation of DWT, Image Text Extraction Technique" *In conference proceedings for Third International Conference on Recent Trends in Computing- 2015 at SRM university NCR India* and published by Procedia Computer Science Journal as special issue (Scopus Indexed): March, 2015