WAVELET BASED TEXTURE CLASSIFICATION OF REMOTLY SENSED IMAGES

A THESIS SUBMITTED TO THE COLLEGE OF SCIENCE, UNIVERSITY OF BAGHDAD IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER IN ASTRONOMY (Remote sensing).

BY

NASR ABID AZIZ AL AZRAQY
B. Sc. (astronomy), 2002

Supervised by

PROFESSOR Dr. SALEH M. ALI AL TOCMACHI

January 2006
بسم الله الرحمن الرحيم

"يعلمُ ما بينَ أيديهم وما خلفهم ولا يحيطون به علما"

صدق الله العظيم

الآية (110) سورة طه
TO MY

PARENTS

AND

FAMILY
The preparation of this thesis was made under our supervision at the College of Science, University of Baghdad in partial fulfillment of the requirements for the Degree of Master in astronomy, (Remote Sensing Application).

Signature:
Title: Professor.

Name: Dr. Saleh M. Ali
Supervisor
Date:

Approved by the Head of department of Astronomy

Signature:
Title: Assistant Professor.

Name: Dr. Ali. Taleb
Head of the Department of Astronomy
Date:
We certify that we have read this thesis which will be submitted to examining committee. In our opinion it is adequate for the partial fulfillment of the requirements for the Degree of Master in Astronomy, (Remote Sensing Application).

Signature:
Title: Assistant Professor.
Name: Dr. Salah A. Saleh
Chairman

Signature:                                            Signature:
Title: Assistant Professor.                       Title: Lecturer.
Name: Dr. Bushra Q. Naqeb               Name: Dr. Alaa S. Mahdi
Member                                               Member

Signature:                                            Signature:
Title: Professor.                                              .
Name: Dr. Saleh M. Ali
Supervisor

Approved by the Dean of College of Science.

Signature:
Title: Professor.
Name: Dr.
The Dean of College of Science
Acknowledgements

I would like to thank my supervisor, Professor. SALEH MAHDI ALI, for this constant help and supervision throughout the course of this work.

I thank the Head of Department of Astronomy at University of Baghdad Dr. Ali. Taleb for his general assistance.

Finally, I shall always be grateful to my friends for their love and great supports in all work steps.
ABSTRACT

In remote sensing, classification methods are usually improved by adopting multi-spectral-bands of satellite images. The problem is arising when only single band is available. In this research, the stationary wavelet transform is adopted to generate multi-band images from the available single band (i.e. the transformed bands have the same size as the original image).

The well known and common classification methods that, usually, followed by remote sensing data users are those categorized as supervised and unsupervised methods. Mostly, these classification techniques either clustering dependent, or statistical-features depending methods.

For expert image processing scientists, it is very well known that entropy features refer to the amount of information within an image. Our present research adopted the relative entropy of the transferred single band into multi-band images to perform the classification. The transformation used in this research is the stationary-wavelet transforms that yield multi-same-size images.

The method is performed on different single band and multi-bands images, using different differentiation block sizes.

On the basis of different wavelet levels and different utilized block sizes, the obtained results are compared with the common supervised classification techniques (i.e. minimum distance and parallelepiped classifiers). The results proved that our presented
method yields, approximately, same results as those obtained by the minimum distance classifier, using multi-bands.

The presented method has been performed by designing a Visual-basic program with a number of routines conducted to perform the classification. Larger blobs of batches representing the image regions have been gained by our introduced method as compared with those obtained by other classification techniques (using ready software package; i.e. ENVI).
## CONTENTS

<table>
<thead>
<tr>
<th>Subject</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chapter One: General Introduction.</strong></td>
<td></td>
</tr>
<tr>
<td>1.1 Introduction.</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Image Data Type.</td>
<td>3</td>
</tr>
<tr>
<td>1.2.1 Binary Images.</td>
<td>4</td>
</tr>
<tr>
<td>1.2.2 Gray-Scale Images</td>
<td>4</td>
</tr>
<tr>
<td>1.2.3 Color Images</td>
<td>4</td>
</tr>
<tr>
<td>1.2.4 Multispectral Image</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Digital Images Representation.</td>
<td>5</td>
</tr>
<tr>
<td>1.4 Common Values.</td>
<td>6</td>
</tr>
<tr>
<td>1.5 The Basic of Remote Sensing.</td>
<td>7</td>
</tr>
<tr>
<td>1.5.1 Remote sensing Model.</td>
<td>8</td>
</tr>
<tr>
<td>1.5.2 Remote sensing Image.</td>
<td>9</td>
</tr>
<tr>
<td>1.5.3 Classification of Satellite Image.</td>
<td>9</td>
</tr>
<tr>
<td>1.6 LANDSAT series.</td>
<td>10</td>
</tr>
<tr>
<td>1.7 Available Data.</td>
<td>11</td>
</tr>
<tr>
<td>1.8 Literature review.</td>
<td>13</td>
</tr>
</tbody>
</table>

| **Chapter Two: The Wavelets Transform.** | |
| 2.1 Discrete Transform. | 17 |
| 2.2 The Wavelet Transform. | 18 |
| 2.3 Wavelet Overview. | 19 |
| 2.4 Historical Perspective. | 21 |
| 2.4.1 Before-1930. | 21 |
| 2.4.2 The 1930S. | 22 |
| 2.4.3 1960-1980. | 23 |
| 2.4.4 Post-1980. | 24 |
| 2.5 What Do Some Wavelet Look Like?. | 24 |
| 2.6 Wavelet Analysis. | 25 |
| 2.6.1 The Discrete Wavelet Transforms. | 25 |
| 2.6.2 The Fast Wavelet Transforms. | 30 |
| 2.6.3 Wavelet Packets. | 30 |
| 2.6.4 Adapted Waveforms. | 30 |
| 2.7 Fourier Transform. | 32 |
| 2.7.1 Fourier Analysis. | 32 |
2.7.2 Discrete Fourier Transform. 33
2.7.3 Windows Fourier Transform. 33
2.7.4 Fast Fourier Transform. 34
2.8 Similarities Between Fourier and Wavelet Transforms. 35
2.9 Dissimilarities Between Fourier and Wavelet Transforms. 35

Chapter Three: Image Classification.
3.1 Introduction. 39
3.2 Simple Pixel Based Classifiers. 39
3.3 Supervised Classification. 40
3.3.1 Minimum-Distance-to-Means. 40
3.3.2 Parallelopiped Classifier. 40
3.3.3 Maximum Likelihood. 40
3.4 Unsupervised Classification. 42
3.4.1 K-Means. 43
3.4.2 Fuzzy C Means. 43
3.4.3 Isodata. 43
3.5 Hyperspectral Classifiers. 43
3.5.1 Band Selection. 43
3.5.2 Spectral Unmixing. 44
3.5.3 Spectral Angle Mapping (SAM). 44
3.6 Texture Classification using DWB. 45
3.6.1 Discriminate Measure. 46

Chapter Four: Results Analysis.
4.1 Introduction. 49
4.2 Adaptive Method Results. 49
4.3 Other classification methods Results. 55
4.4 Compression between the Methods  60

Chapter Five: Conclusions & Suggestion
5.1 Conclusions. 64
5.2 Suggestions for Future Works. 66
<table>
<thead>
<tr>
<th>Figure No.</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1-1)</td>
<td>Illustrate an image example.</td>
<td>6</td>
</tr>
<tr>
<td>(2-1)</td>
<td>The fractal self-similarity of the Daubechies</td>
<td>24</td>
</tr>
<tr>
<td>(2-2)</td>
<td>Several different families of wavelets.</td>
<td>25</td>
</tr>
<tr>
<td>(2-3)</td>
<td>(a-b-c) Original image, 1st and 2nd level of Wavelet transform.</td>
<td>28</td>
</tr>
<tr>
<td>(2-4)</td>
<td>Fourier basis functions, time-frequency tiles, and coverage of the time-frequency plane.</td>
<td>36</td>
</tr>
<tr>
<td>(2-5)</td>
<td>Daubechies wavelet basis functions, time-frequency tiles, and coverage of the time-frequency plane.</td>
<td>37</td>
</tr>
<tr>
<td>(3-1)</td>
<td>Minimum-Distance Classifier.</td>
<td>41</td>
</tr>
<tr>
<td>(3-2)</td>
<td>Parallelopiped Classifier</td>
<td>42</td>
</tr>
<tr>
<td>(3-3)</td>
<td>Maximum Likelihood Classifier.</td>
<td>42</td>
</tr>
<tr>
<td>(3-4)</td>
<td>Spectral Angle Mapping (SAM).</td>
<td>45</td>
</tr>
<tr>
<td>(4-1)</td>
<td>Demonstration of the classification results with different adopted bands (i.e. transformation levels) and block's sizes.</td>
<td>53</td>
</tr>
<tr>
<td>(4-2)</td>
<td>Illustrating the relationship between number of classified points within each class, using different block's sizes (i.e. C1=3, C2=5, and C3=7), for wavelet transform of one-level.</td>
<td>53</td>
</tr>
<tr>
<td>(4-3)</td>
<td>Illustrating the relationship between classified points within each class with different levels of wavelet transformation (i.e. C1=level-one, C4=level-two, and C7=level-three), using fixed block size=3.</td>
<td>54</td>
</tr>
<tr>
<td>(4-4)</td>
<td>Same as in figure (4-10), but by adopting block size=5.</td>
<td>54</td>
</tr>
<tr>
<td>(4-5)</td>
<td>Same as above with block size=7.</td>
<td>55</td>
</tr>
<tr>
<td>(4-6)</td>
<td>Classification results by minimum distance classifier, using single-band image.</td>
<td>56</td>
</tr>
<tr>
<td>(4-7)</td>
<td>Classification result obtained by performing the supervised parallelepiped algorithm, using single-band image.</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>(4-8)</td>
<td>Minimum distance classification result, using 6-bands.</td>
<td></td>
</tr>
<tr>
<td>(4-9)</td>
<td>Parallelepiped classification result, using 6-bands.</td>
<td></td>
</tr>
<tr>
<td>(4-10)</td>
<td>Comparison between C3 (figure 4-7), and the minimum distance classifier (one-level wavelet transformation, and block size=7).</td>
<td></td>
</tr>
<tr>
<td>(4-11)</td>
<td>Comparison between C3 (figure 4-7), and the parallelepiped classifier (one-level wavelet transformation, and block size=7).</td>
<td></td>
</tr>
<tr>
<td>(4-12)</td>
<td>Comparison between C3 (figure 4-7), and the minimum distance classifier (one-level wavelet transformation, and block size=7), implemented on 6-band images.</td>
<td></td>
</tr>
<tr>
<td>(4-13)</td>
<td>Comparison between C3 (figure 4-7), and the parallelepiped classifier (one-level wavelet transformation, and block size=7), implemented on 6-band images.</td>
<td></td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table No.</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1-1)</td>
<td>Types of image operations.</td>
<td>5</td>
</tr>
<tr>
<td>(1-2)</td>
<td>Common values of digital image parameters.</td>
<td>6</td>
</tr>
<tr>
<td>(1-3)</td>
<td>Thematic Mapper Spectral Bands.</td>
<td>12</td>
</tr>
<tr>
<td>(4-1)</td>
<td>representing the classified number of pixels for each region, using different block's sizes.</td>
<td>52</td>
</tr>
<tr>
<td>(4-2)</td>
<td>represents the number of points and the decided color for each class.</td>
<td>56</td>
</tr>
<tr>
<td>(4-3)</td>
<td>represents the number of points and the decided color for each class.</td>
<td>57</td>
</tr>
<tr>
<td>(4-4)</td>
<td>represents the number of points and the decided color for each class, using the minimum distance classifier on 6-bands.</td>
<td>58</td>
</tr>
<tr>
<td>(4-5)</td>
<td>represents the number of points and the decided color for each class, using the parallelepiped classifier on 6-bands</td>
<td>59</td>
</tr>
</tbody>
</table>
GLOSSARY

DFT  Discrete Fourier Transform.
DN   Determines the Mean value.
DWB  Discrete Wavelet packet
DWT  Discrete Wavelet Transform.
ENVI ENvironmental Visualization Image (Software package)
ESC  Export System Classifier.
FFT  Fast Fourier Transform.
MESMA Multiple Endmember Spectral Mixture Analysis.
Landsat Earth Resources Technology Satellite.
MLC  Maximum Likelihood Classifier.
NNC  Neural Network Classifier.
ROI  Region of Interest.
RS   Remote Sensing.
SAM  Spectral Angle Mapping.
TM   Thematic Mapper, sensor on board Landsat 4, and 5.
WFT  Windowed Fourier Transform.
CHAPTER ONE

GENERAL INTRODUCTION
1.1 Introduction

Digital image processing is an area, which has been widely investigated and there is abundant literature covering different aspects of processing images. Two major application areas motivate the interest in digital image processing methods: improvement of pictorial information for the human interpretation and data processing for autonomous machine perception [1].

The processing methods can be classified into several categories.

(a) Image transforms. The Fourier transform is by far the most popular, but there are several two-dimensional operators who can be applied to images for enhancement, restoration, encoding, and description purposes.

(b) The objective of image enhancement techniques is to process an image so that the result is more suitable than the original image for a specific application. The techniques in this category are very much problem-oriented and they fall into two broad subcategories: spatial domain techniques, which directly manipulate the pixels in an image, and frequency domain techniques, based on modifying the Fourier transform of an image.

(c) Image restoration techniques are similar to image enhancement techniques in the sense that their ultimate goal is also to improve an image in some sense. As opposed to image enhancement methods, which are in essence heuristic procedures designed to manipulate an image to take advantage of the human visual system, restoration methods attempt
to reconstruct an image that has been degraded by using some knowledge of the degradation phenomenon.

(d) Image compression addresses the problem of reducing the amount of data required to represent a digital image. The reduction is based on removal of redundant data.

(e) Image segmentation is usually the first step in image analysis. Methods in this category attempt to subdivide an image into its components up to a level, which depends on the problem being solved.

(f) Representation and description techniques are used to represent and describe aggregates of segmented pixels in a form suitable for computer processing.

(g) Recognition and interpretation methods are especially suitable for applications requiring automated image analysis and usually involve some degree of artificial intelligence.

1.2 Image Data Type

The digital image \( f(x, y) \) is represented as two-dimensional array of data, where each pixel value corresponds to the brightness of the image at the point \( (x, y) \). The simplest type of image is the monochrome (one color, this is what we normally refer to as gray image data). Other types of image data required extension or modification to this model, typically there are multi band images (color, multispectral), and they can be modeled by using different functions \( f(x, y) \) each corresponding to single
separate band of brightness information. The image can be classified into the following types [1].

1.2.1 Binary Images
Binary images are the simplest type of images where each image element can take one of two values; typically black and white (i.e., ‘0’ and ‘1’).

1.2.2 Gray-Scale Images
Gray-scale images are referred to as monochrome or one color images, they contain brightness information only, no color information (i.e., 0…255).

1.2.3 Color Images
Color images can be modeled as three band monochrome image data, where each band of data corresponds to different color. The actual information stored in the digital image is the brightness information in each spectral band. When the image is displayed, the corresponding brightness information is displayed on the screen by picture elements that emit light energy corresponding to that particular color. Typical color images are represented as red, green, and blue or RGB images. Using the 8-bits monochrome standard as a model, the corresponding color image would have 24-bits/pixel-8-bits for each of the three-color bands (red, green, and blue).

1.2.4 Multispectral Images
To extract earth information, a technique of “remote sensing” is used, it means the observation of, or gathering information about a target by a sensor separated from it by some distance. This information is represented as Multispectral images; they are not image in usual sense
because some of the information is not directly visible by the human system.

1.3 Digital Images Representation

There are many ways to classifying and characterize image operations. The reason for doing so is to understand what type of results we might expect to achieve with a given type of operation or what might be the computational burden associated with a given operation.

The types of operations that can be applied to digital images to transform an input image \(a[m, n]\) into an output image \(b[m, n]\) (or another representation) can be classified into three categories as shown in Table (1-1).

<table>
<thead>
<tr>
<th>Operation</th>
<th>Characterization</th>
<th>Generic Complexity/Pixel</th>
</tr>
</thead>
<tbody>
<tr>
<td>* Point</td>
<td>- The output value at a specific coordinate is dependent only on the input value at that same coordinate.</td>
<td>(\text{constant})</td>
</tr>
<tr>
<td>* Local</td>
<td>- The output value at a specific coordinate is dependent on the input values in the neighborhood of that same coordinate.</td>
<td>(P^2)</td>
</tr>
<tr>
<td>* Global</td>
<td>- The output value at a specific coordinate is dependent on all the values in the input image.</td>
<td>(N^2)</td>
</tr>
</tbody>
</table>
1.4 Common Values

There are standard values for the various parameters encountered in digital image processing. These values can be caused by video standards, by algorithmic requirements, or by the desire to keep digital circuitry simple. Table (2-1) shows some commonly encountered values, where figure (1-1) illustrates an image example.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Typical values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>$N$</td>
<td>256, 512, 525, 625, 1024, 1035</td>
</tr>
<tr>
<td>Columns</td>
<td>$M$</td>
<td>256, 512, 768, 1024, 1320</td>
</tr>
<tr>
<td>Gray Levels</td>
<td>$L$</td>
<td>2, 64, 256, 1024, 4096, 16384</td>
</tr>
</tbody>
</table>

Figures (1-1) illustrate an image example.

The number of distinct gray levels is usually a power of 2, that is, $L=2^B$ where $B$ is the number of bits in the binary representation of the brightness levels. When $B>1$ we speak of a gray-level image; when $B=1$ we speak of a binary image. In a binary image there are just two gray levels which can be referred to, for example, as "black" and "white" or "0" and "1".
1.5 The Basis of Remote Sensing

Remote sensing can be defined as “art and science of acquiring information about targets without coming into direct contact between sensor and target”, [2]. The definition holds two main aspects, **Data Collection**, and **Data Analysis**. This will be done by sensing, recording reflected or emitted energy, processing, analyzing, and applying that information.

After processing, image is interpreted, visually and / or digitally or electronically, to extract information about the ground target which was illuminated. Extraction of information about round cover from the satellite imagery is called interpretation of digital image data in the field of satellite remote sensing. Interpretation of an image can be done in two ways, quantitative analysis and photo-interpretation. The former involves the use of a computing device which examine each pixel in the image on the basis of certain characteristics of that pixel and take decision about the pixel type and finally by counting number of such pixels in the image area covered by those pixels can estimated. The later one use human analyst or interpreter who gets information by visual inspection of imagery.

Remote sensing system can be detecting the variations in the **Spatial**, **Spectral**, and **Temporal** in the real world. The first one is the target dimension, the second is the color and reflected energy variations, where the last is the variations among the time, [3].

The main toll in the remote sensing system is the sensed devices, called “Sensor”. The most sensor types are works in the space (outer of
atmosphere) with different spatial and spectral resolution, and different spectral bands, [2].

The new advantages of this technique can be summarized according to as following [4]:

1. Synoptic view.
2. Multi spectral, i.e. many spectral images for one scene.
3. Repeatable capability, i.e. exposure the same areas after fixed time.
4. Sun-Synchronous, (exposure any position in the world in the same daytime).
5. Low cost and saves of time.
6. The output formula is in digital form, (easy in treatment and copy and reprocessing, also many mathematical and statistical operations can be applied.)

1.5.1 Remote Sensing Model

The remote sensing model consists of four major components, Target, Sensor, Radiation Source, and Transmission Path [5]. The targets are any features or phenomenon on the Earth surface or atmosphere, where, the sensor are the devices that records the reflected or emitted radiation from the target. Sensors can be divided according to energy source as Active & Passive. The sensor system parameters play an important role in sensing process, such as Spatial resolution, and Spectral resolution. The process of remote sensing depend on the reflected or emitted energy from the target and background, this energy can be supplied from the sun as major part of energy. The useful wavelengths of sun radiation are (0.3 µm to 100 cm). The transmission path affected the remote sensing results accuracy by much process such as Scattering, Absorption and Diffraction.
1.5.2 The Remote Sensing Image
The output results of many remote sensing processes are digital images, which are often multispectral bands. Many essential processes must be applied on an image in order to be useful. These processes include preprocessing, enhancement, and classification and layers extraction. There are multiple classification algorithms based on Supervised and Unsupervised methods. Our aim is to modify a classification method which is applied for wavelet transform results and use new criteria called Killback-Leber Distance on supervised method in order to extract the image features.

1.5.3 Classification of Satellite Image
The classification of satellite image is a prime component of any quantitative analysis process [6]. The process of classification involves labeling of each pixel of image into a ground cover type using its numeric values in different spectral channels. In other words, image classification is the process of creating thematic maps from satellite imagery, a thematic map is an informational representation of an image which shows the spatial distribution of a particular class. Classes can be as diversified as their areas of interest for example soil, vegetation, water, and clouds. Inside a class, there can be defined subclasses, and thus the process of classification needs to become more refined. The mathematical technique of pattern recognition is used to classify each pixel of the image corresponding to specific ground cover types. Here the pattern means a pixel itself or a vector which has elements from all available spectral bands in the form of brightness values arranged in column form. Supervised classification and unsupervised classification are the two broad types of classification processes that are used in satellite remote sensing [6].
1.6 LANDSAT Series.

Landsat is unmanned system that prior to 1974 was called ERTS (Earth Resources Technology Satellite). The system operates in the international public domain, which means that, open Sky, NASA Godard and EROS data center archives all images, and user in the would may purchase all images at uniform prices and priorities. Landsat program is largely responsible for the growth and acceptance or Remote Sensing as a scientific discipline. Landsat satellite have been placed in orbit using Delta rockets launched from Vandenberg Air Force Base on California. The three satellites in this series were launched July 23; 1972, January 21; 1975, and March 5; 1978. The sensors on board these satellite are RBV (Return Beam Vidicon Camera) and MSS (Multi Spectral Scanner). All have ceased operation, but they produced hundreds of thousand of valuable images. The second generation of Landsat consist of two satellites launched July 16; 1982, and March 1; 1984. The powerful Earth imaging sensor is TM (Thematic Mapper). After more than two decades of success, the Landsat program realized its first unsuccessful mission with the October; 1993, launch failure of Landsat 6. Landsat 7 was launched on April 15; 1999. The Earth observing instrument on board this spacecraft is the ETM+ (Enhance Thematic Mapper Plus). Table (A-1) illustrates the main sensor parameters on board Landsat series. Landsat satellites have the same orbit characteristics such as: circular near polar, sawth to sawth west in direction, sun-synchronous, synoptic view and repeatable capability. Date gathered by the scanners in cross truck direction. The gathered data stored on board satellite and then transmit to Earth station by use substation in the space, (DOMSAT) satellite. The data format is variable for sensors to another, as example, MSS images has 7-bits representation, TM images has 8-bits representation, and
ETM+ images have 8 and 9 bits representation. Also, these sensors work in multi spectral bands in the electromagnetic spectrum.

1.7 Available Data

The study area located in west of IRAQ, Habania aria, geographic coordinates “Latitude (33º 00' 15") to (33º 50' 00") N, Longitude (43º 03' 45") to (44º 15' 00") E. The general land cover on the area are Urban and the habania lake, vegetation cover is around lake area, many geological features can be signed such as, and rock erosion. The relief highest is (75 m) above sea level.

The available data consist of multibands LANDSAT-5 THEMATIC MAPPER TM images, Image size is 256 in rows and 256 in columns.
Table (1-3) Thematic Mapper Spectral Bands, [2]

<table>
<thead>
<tr>
<th>Band</th>
<th>Wavelength (micro-m)</th>
<th>Nominal spectra location</th>
<th>Principal applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.45-0.52</td>
<td>Blue</td>
<td>Designed for water body penetration, making it useful for coastal water mapping. Also useful for soil/vegetation discrimination, forest type mapping, and cultural feature identification.</td>
</tr>
<tr>
<td>2</td>
<td>0.52-0.60</td>
<td>Green</td>
<td>Designed to measure green reflectance peak of vegetation for vegetation discrimination and vigor assessment. Also useful for cultural feature identification.</td>
</tr>
<tr>
<td>3</td>
<td>0.63-0.69</td>
<td>Red</td>
<td>Red Designed to sense in a chlorophyll absorption region aiding in plant species differentiation. Also useful for cultural feature identification.</td>
</tr>
<tr>
<td>4</td>
<td>0.76-0.90</td>
<td>Near-infrared</td>
<td>Useful for determining vegetation types, vigor, and biomass content, for delineating water bodies, and for soil moisture discrimination.</td>
</tr>
<tr>
<td>5</td>
<td>1.55-1.75</td>
<td>Mid-infrared</td>
<td>Indication of vegetation moisture content and soil moisture. Also useful for differentiation of snow from clouds.</td>
</tr>
<tr>
<td>6</td>
<td>10.4-12.5</td>
<td>Thermal infrared</td>
<td>Useful in vegetation stress analysis, soil moisture discrimination, and thermal mapping applications.</td>
</tr>
<tr>
<td>7</td>
<td>2.08-2.35</td>
<td>Mid-infrared</td>
<td>Useful for discrimination of mineral and rock types. Also sensitive to vegetation moisture content.</td>
</tr>
</tbody>
</table>
1.8 Literature Review

Many works have been shown that different classification algorithms produce different results, specially between the widely applied traditional maximum likelihood classifier (MLC) and the neural network classifier (NNC) [Hepner et al., 1990, Benediktsson et al., 1990, Key et al. 1990, Bischof et al. 1992, Kanellopoulos et al. 1992, Civco 1993, Paola and Schowengerdt, 1994, Solaiman and Mouchot 1994, Skidmore et al., 1997]. It is often found in practical application that for some classes a neural classifier gives the best results whilst for other classes a statistical classifiers gives the best results [7].


[Kettig and Landgrebe, 1976], proposed a classification method to isolate homogeneous objects from multi-spectral satellite images. [12]

[Skidmore, 1989, Benediktsson et al., 1990], study the minimum Mahalanobis distance classifier, the K-means, and K-nearest neighbor clustering methods. [9]

[Saito,T., Higuchi, H., and Komatsu,T.,1991]. Utilized a multi-resolution technique in computer vision for tasks such as object recognition and motion estimation as well as in image compression, with sub band coding. [13]

[Arai, 1992], introduced a new approach that A purification of training samples in the supervised classification of TM – images. [14]
Leo, R.S., Jung, K., Anderson, K., Shen, S., Lawton, W. 1992, investigate multi classification methods, they found that the Wavelet band features are used to classify transient acoustic signals produced by impulses.[15]

Muneychika, 1993, improved the classification accuracy of multispectral images by using some resolution enhancement techniques. [16]

Sari-Sarraf, Hamed, 1993, The theoretical development of a novel technique for the representation of one and two-dimensional discrete signals, called the multi scale wavelet representation. [17]

Solaiman and Mouchot, Fierens et al. 1994, stated that the reasons for the differences in classification accuracy are not completely understood. [10]

Vetterii, M., 1995, has proven that the wavelet transform plays substantial role in multi-resolution technique. [18]

Brown et al., 1998, applied NNC and MLC on supervised classification of glaciated landscapes. NNC reproduced better the patterns of the less a really extensive, whereas the overall pattern of all classes was better reproduced with the MLC. Some researchers also involved other classifiers to be compared, like the expert system classifier (ESC) (Wilkinson et al. 1992, Skidmore 1989). [8]

Punya Thitimajshima, 1999, presented a novel method for SAR image classification based on stationary wavelet transform. [19]
[Al-Obaidy, H.A., 2000], has shown that the application of wavelets image compression has drawn more interest than any other earlier applications. [20]
CHAPTER TWO

THE WAVELET TRANSFORM
Chapter Two
The Wavelet Transform

2.1 Discrete Transforms

The transforms considered here provide information regarding the spatial frequency content of an image. In general, a transform maps image data into a different mathematical space via a transformation equation, where the transform maps image from the spatial domain to the frequency domain, [21].

These transforms are used as tools in many areas of engineering and science, including computer imaging. Originally defined in their continuous forms, they are commonly used today in their discrete forms, where they can be processed in digital computers in a more appropriate way [22]. The large number of arithmetic operations required for these discrete transforms, combined with the massive amounts of data in an image, requires a great deal of computer power. The ever-increasing compute power, memory capacity, and disk storage available today make the use of these transforms much more feasible than in recent years, [21].

The ways in which the image brightness levels change in space define the spatial frequency. For example, rapidly changing brightness corresponds to high spatial frequency, whereas slowly changing brightness levels relate to low-frequency information. The lowest spatial frequency, called the zero frequency term, corresponds to an image with a constant value.

The discrete form of these transforms is created by sampling the continuous form of the function on which these transforms are based; measuring the value of the function at discrete intervals in space, that is,
the basis junctions. The sampling process, for the one-dimensional case, provides us with basis vectors. When extending these into two-dimensions, they are basis matrices or basis images, [23].

The general form of the transformation equation, assuming an image with $N \times N$ size, is given as follow [21]:

$$T(u, v) = \sum_{r=0}^{N-1} \sum_{c=0}^{N-1} I(r, c)B(r, c; u, v) \quad (2-1)$$

Hence, $I(r, c)$ is the two-dimensional image, $u$ and $v$ are the frequency domain variables. $T(u, v)$ is the transform coefficient, and $B(r, c; u, v)$ corresponds to basis images, corresponding to each different value for $u$ and $v$, and the size of each is $r$ by $c$.

To obtain the image from the transform coefficients, we apply the inverse transform equation:

$$I(r, c) = T^{-1}[T(u, v)] = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} T(u, v) B^{-1}(r, c; u, v) \quad (2-2)$$

Where $T^{-1}[T(u, v)]$ represents the inverse transform, and the $B^{-1}(r,u;u,v)$, represents me inverse basis images.

**2.2 THE WAVELET TRANSFORM**

Wavelets can be described as a transform that has basis functions that are shifted and expanded versions of themselves. Because of this, the wavelet transform contains not just frequency information but spatial information as well [25]. Wavelets were developed independently in the
fields of mathematics, quantum physics, electrical engineering, and seismic geology. Interchanges between these fields during the last ten years have led to many new wavelet applications such as image compression, turbulence, human vision, radar, and earthquake prediction. The history of wavelets will be described beginning with Fourier transform, compare wavelet transforms with Fourier transforms, state properties and other special aspects of wavelets.

2.3 Wavelet Overview

The fundamental idea behind wavelets is to analyze according to scale. Indeed, some researchers in the wavelet field feel that, by using wavelets, one is adopting a whole new mindset or perspective in processing data.

Wavelets are functions that satisfy certain mathematical requirements and are used in representing data or other functions. This idea is not new. Approximation using superposition of functions has existed since the early 1800's, when Joseph Fourier discovered that he could superpose sines and cosines to represent other functions. However, in wavelet analysis, the scale that we use to look at data plays a special role. Wavelet algorithms process data at different scales or resolutions. If we look at a signal with a large window," we would notice gross features. Similarly, if we look at a signal with a small window," we would notice small features. The result in wavelet analysis is to see both the forest and the trees, so to speak.

This makes wavelets interesting and useful. For many decades, scientists have wanted more appropriate functions than the sines and cosines, which comprise the bases of Fourier analysis, to approximate choppy signals [26]. By their definition, these functions are non-local (and stretch out to infinity). They therefore do a very poor job in approximating sharp
spikes. But with wavelet analysis, we can use approximating functions that are contained neatly in finite domains. Wavelets are well suited for approximating data with sharp discontinuities.

The wavelet analysis procedure is to adopt a wavelet prototype function, called an analyzing wavelet or mother wavelet. Temporal analysis is performed with a contracted, high frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low frequency version of the same wavelet. Because the original signal or function can be represented in terms of a wavelet expansion (using coefficients in a linear combination of the wavelet functions), data operations can be performed using just the corresponding wavelet coefficients. And if you further choose the best wavelets adapted to your data, or truncate the coefficients below a threshold, your data is sparsely represented. This sparse coding makes wavelets an excellent tool in the field of data compression, [21].

Other applied fields that are making use of wavelets include astronomy, acoustics, nuclear engineering, sub-band coding, signal and image processing, neurophysiology, music, magnetic resonance imaging, speech discrimination, optics, fractals, turbulence, earthquake-prediction, radar, human vision, and pure mathematics applications such as solving partial differential equations.
2.4 HISTORICAL PERSPECTIVE

In the history of mathematics, wavelet analysis shows many different origins [27]. Much of the work was performed in the 1930s, and, at the time, the separate efforts did not appear to be parts of a coherent theory.

2.4.1 Before -1930

Before 1930, the main branch of mathematics leading to wavelets began with Joseph Fourier (1807) with his theories of frequency analysis, now often referred to as Fourier synthesis. He asserted that any 2-periodic function $f(x)$ is given the summation, equation (2-1);

$$a_o + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \tag{2-3}$$

of its Fourier series. The coefficients $a_o$, $a_k$ and $b_k$ are calculated by equations (2-2);

$$a_o = \frac{1}{2\pi} \int_{0}^{2\pi} f(x)dx \quad a_k = \frac{1}{\pi} \int_{0}^{2\pi} f(x)\cos(kx)dx \quad b_k = \frac{1}{\pi} \int_{0}^{2\pi} f(x)\sin(kx)dx \tag{2-4}$$

Fourier's assertion played an essential role in the evolution of the ideas mathematicians had about the functions. He opened up the door to a new functional universe.

After 1807, by exploring the meaning of functions, Fourier series convergence, and orthogonal systems, mathematicians gradually were led from their previous notion of frequency analysis to the notion of scale analysis. That is, analyzing $f(x)$ by creating mathematical structures that vary in scale. By Construct a function, shift it by some amount, and change its scale. Apply that structure in approximating a signal. Now
repeat the procedure. Take that basic structure, shift it, and scale it again. Apply it to the same signal to get a new approximation. And so on. It turns out that this sort of scale analysis is less sensitive to noise because it measures the average fluctuations of the signal at different scales.

The first mention of wavelets appeared in an appendix to the thesis of A. Haar (1909). One property of the Haar wavelet is that it has compact support, which means that it vanishes outside of a finite interval. Unfortunately, Haar wavelets are not continuously differentiable which somewhat limits their applications.

2.4.2 The 1930s

In the 1930s, several groups working independently researched the representation of functions using scale-varying basis functions. Understanding the concepts of basis functions and scale-varying basis functions is key to understanding wavelets; the sidebar below provides a short detour lesson for those interested.

By using a scale-varying basis function called the Haar basis function (more on this later) Paul Levy, a 1930s physicist, investigated Brownian motion, a type of random signal [27]. He found the Haar basis function superior to the Fourier basis functions for studying small complicated details in the Brownian motion.

Another 1930s research effort by Littlewood, Paley, and Stein involved computing the energy of a function f(x):

\[
\text{energy} = \frac{1}{2} \int_0^{2\pi} |f(x)|^2 \, dx
\]  

(2-5)

The computation produced yield different results if the energy was concentrated around a few points or distributed over a larger interval. This result disturbed the scientists because it indicated that energy might
not be conserved. The researchers discovered a function that can vary in scale and can conserve energy when computing the functional energy. Their work provided David Marr with an effective algorithm for numerical image processing using wavelets in the early 1980s.

2.4.3 1960-1980

Between 1960 and 1980, the mathematicians Guido Weiss and Ronald R. Coifman studied the simplest elements of a function space, called atoms, with the goal of finding the atoms for a common function and finding the assembly rules" that allow the reconstruction of all the elements of the function space using these atoms. In 1980, Grossman and Morlet, a physicist and an engineer, broadly defined wavelets in the context of quantum physics. These two researchers provided a way of thinking for wavelets based on physical intuition.

2.4.4 Post-1980

In 1985, Stephane Mallat gave wavelets an additional jump-start through his work in digital signal processing. He discovered some relationships between quadrature mirror filters, pyramid algorithms, and orthonormal wavelet bases. Inspired in part by these results, Y. Meyer constructed the first non-trivial wavelets. Unlike the Haar wavelets, the Meyer wavelets are continuously differentiable; however they do not have compact support. A couple of years later, Ingrid Daubechies used Mallat's work to construct a set of wavelet orthonormal basis functions that are perhaps the most elegant, and have become the cornerstone of wavelet applications today.
2.5 WHAT DO SOME WAVELETS LOOK LIKE?

Wavelet transforms comprise an infinite set. The different wavelet families make different trade-offs between how compactly the basis functions are localized in space and how smooth they are.

Some of the wavelet bases have fractal structure. The Daubechies wavelet family is one example, (see Figure 2-3).

![Figure 2-1](image)

**Figure (2-1), The fractal self-similarity of the Daubechies**

Within each family of wavelets (such as the Daubechies family) are wavelet subclasses distinguished by the number of coefficients and by the level of iteration. Wavelets are classified within a family most often by the number of vanishing moments. This is an extra set of mathematical
Figure (2-2) Several different families of wavelets

Relationships for the coefficients that must be satisfied, and is directly related to the number of coefficients [26]. For example, within the Coiflet wavelet family are Coiflets with two vanishing moments, and Coiflets with three vanishing moments. In Figure (2-4), I illustrate several different wavelet families.

2.6 WAVELET ANALYSIS

Now we begin our tour of wavelet theory, when we analyze our signal in time for its frequency content. Unlike Fourier analysis, in which we analyze signals using sines and cosines, now we use wavelet functions.

2.6.1 THE DISCRETE WAVELET TRANSFORM

Dilations and translations of the "Mother function," or analyzing wavelet" $\Phi(x)$; define an orthogonal basis, our wavelet basis:

$$\Phi_{(s,t)}(x) = 2^{s/2} \Phi(2^{-3} x - t)$$
The variable s and l are integers that scale and dilate the mother function $\Phi$ to generate wavelets, such as a Daubechies wavelet family. The scale index s indicates the wavelet's width. That the mother functions are rescaled, or “dilated” by powers of two, and translated by integers. What makes wavelet bases especially interesting is the self-similarity caused by the scales and dilations.

To span our data domain at different resolutions, the analyzing wavelet is used a scaling equation;

$$W(x) = \sum_{k=1}^{N-2} (-1)^k c_{k+1} \Phi(2x + k)$$  \hspace{1cm} (2-7)

Where $W(x)$ is the scaling function for the mother function $\Phi$; and $c_k$ are the wavelet coefficients The wavelet coefficients must satisfy linear and quadratic constraints of the form;

$$\sum_{k=0}^{N-1} c_k = 2 \quad \sum_{k=0}^{N-1} c_k c_{k+2i} = 2 \delta_{i,0}$$  \hspace{1cm} (2-8)

Where, $\delta$ is the delta function and i is the location index

One of the most useful features of wavelets is the ease with which a scientist can choose the defining coefficients for a given wavelet system to be adapted for a given problem. In Daubechies original paper [31], she developed specific families of wavelet systems that were very good for representing polynomial behavior. The Haar wavelet is even simpler, and it is often used for educational purposes.

Numerous filters can be used to implement the wavelet transform, and two of the commonly used ones, the Daubechies and Haar. These are separable, so they can be used to implement a wavelet transform by first
convolving them with the rows and then the columns. The Haar vectors are simple;

LOWPASS: \( \frac{1}{\sqrt{2}} [1 \ 1] \)

HIGHPASS: \( \frac{1}{\sqrt{2}} [1 \ -1] \)

An example of Daubechies basis vectors follows:

LOWPASS: \( \frac{1}{4\sqrt{2}} [1+\sqrt{3}, \ 3+\sqrt{3}, \ 3-\sqrt{3}, \ 1-\sqrt{3}] \)

HIGHPASS: \( \frac{1}{4\sqrt{2}} [1-\sqrt{3}, \ \sqrt{3}-3, \ 3+\sqrt{3}, \ -1-\sqrt{3}] \)

The wavelet transform is performed by doing the following:

1. Convolve the lowpass filter with the rows (by sliding, multiplying coincident terms, and summing the results).

2. Convolve the lowpass filter with the columns (of the results from step 1) and subsample this result by taking every other value; this gives us the lowpass-lowpass version of the image.

3. Convolve the result from step 1, the lowpass filtered rows, with the highpass filter on the columns. Subsample by taking every other value to produce the lowpass-highpass image.

4. Convolve the original image with the highpass filter on the rows and save the result.

5. Convolve the result from step 4 with the lowpass filter on the columns; subsample to yield the highpass-lowpass version of the image.

6. To obtain the highpass-highpass version, convolve the columns of the result from step 4 with the highpass filter.
The convention for displaying the wavelet transform results, as an image, is shown in figure (2-5) (a-b-c). In figure (2-5) b the results of applying the wavelet transform to an image see the lowpass-lowpass image in the upper-left corner, the lowpass-highpass images on the diagonals, and the highpass-highpass in the lower-right corner. If continue to run the same wavelet transform on the lowpass-lowpass version of the image to get more subimages.

![Figure (2-3 a-b-c) Original image, 1st and 2nd level of Wavelet transform.](image)

The inverse wavelet transform is performed by enlarging the wavelet transform data to its original size. Insert zeros between each of the four subimages, and sum the results to obtain the original image. For the Harr filter, the inverse wavelet filters are identical to the forward filters; for the Daubechies the inverse wavelet filter are:
LOWPASS<sub>inv</sub>: \( \frac{1}{4\sqrt{2}} [3-\sqrt{3}, 3+\sqrt{3}, 1+\sqrt{3}, 1-\sqrt{3}] \)

HIGHPASS<sub>inv</sub>: \( \frac{1}{4\sqrt{2}} [1-\sqrt{3}, -1-\sqrt{3}, 3+\sqrt{3}, -3+\sqrt{3}] \)

The filter or coefficients are placed in a transformation matrix, which is applied to a raw data vector. The coefficients are ordered using two dominant patterns, one that works as a smoothing filter (like a moving average), and one pattern that works to bring out the data's "detail" information. These two orderings of the coefficients are called a quadrature mirror filter pair in signal processing parlance [29].

To complete our discussion of the DWT, let's look at how the wavelet coefficient matrix is applied to the data vector. The matrix is applied in a hierarchical algorithm, sometimes called a pyramidal algorithm. The wavelet coefficients are arranged so that odd rows contain an ordering of wavelet coefficients that act as the smoothing filter, and the even rows contain an ordering of wavelet coefficients with different signs that act to bring out the data's detail. The matrix is first applied to the original, full-length vector. Then the vector is smoothed and decimated by half and the matrix is applied again. Then the smoothed, halved vector is smoothed, and halved again, and the matrix applied once more. This process continues until a trivial number of “smooth-smooth-smooth...” data remain. That is, each matrix application brings out a higher resolution of the data while at the same time smoothing the remaining data. The output of the DWT consists of the remaining smooth (etc.)" components, and all of the accumulated detail" components.
2.6.2 THE FAST WAVELET TRANSFORM

The DWT matrix is not sparse in general, so we face the same complexity issues that we had previously faced for the discrete Fourier transform [32]. We solve it as we did for the FFT, by factoring the DWT into a product of a few sparse matrices using self-similarity properties. The result is an algorithm that requires only order $n$ operations to transform an $n$-sample vector. This is the fast DWT of Mallat and Daubechies.

2.6.3 WAVELET PACKETS

The wavelet transform is actually a subset of a far more versatile transform, the wavelet packet transform [33].

Wavelet packets are particular linear combinations of wavelets [32]. They form bases which retain many of the orthogonality, smoothness, and localization properties of their parent wavelets. The coefficients in the linear combinations are computed by a recursive algorithm making each newly computed wavelet packet coefficient sequence the root of its own analysis tree.

2.6.4 ADAPTED WAVEFORMS

Because we have a choice among an infinite set of basis functions, we may wish to find the best basis function for a given representation of a signal [32]. A basis of adapted waveform is the best basis function for a given signal representation. The chosen basis carries substantial information about the signal, and if the basis description is efficient (that is, very few terms in the expansion are needed to represent the signal), then that signal information has been compressed.
According to Wickerhauser, some desirable properties for adapted wavelet bases are [32]:

1. Speedy computation of inner products with the other basis functions.

2. Speedy superposition of the basis functions.

3. Good spatial localization, so researchers can identify the position of a signal that is contributing a large component.

4. Good frequency localization, so researchers can identify signal oscillations.

5. Independence, so that not too many basis elements match the same portion of the signal.

For adapted waveform analysis, researchers seek a basis in which the coefficients, when rearranged in decreasing order, decrease as rapidly as possible. To measure rates of decrease, they use tools from classical harmonic analysis including calculation of information cost functions. This is defined as the expense of storing the chosen representation. Examples of such functions include the number above a threshold, concentration, entropy, and logarithm of energy, Gauss-Markov calculations, and the theoretical dimension of a sequence.
2.7 Fourier Transform

The Fourier transform is the most well known transform. It was developed by Baptiste Joseph Fourier (1768-1830) to explain the distribution of temperature and heat condition. Since that time Fourier transform has found numerous uses, including vibration analysis in mechanical engineering, circuit analysis in electrical engineering, and in computer imagining. This transform allows for the decomposition of an image into a weighted sum of 2-D sinusoidal terms. Assuming an \( N \times N \) image, the equation for the 2-D discrete Fourier transform is, [24]:

\[
F(u, v) = \frac{1}{N} \sum_{r=0}^{N-1} \sum_{c=0}^{N-1} I(r, c) e^{-j \frac{2 \pi (ur + vc)}{N}}
\]

(2-9)

The base of the natural logarithm function \( e \) is about 2.71828; \( j \), the imaginary coordinate for a complex number, equals \( \sqrt{-1} \)

To get our original image back, we need to apply the inverse transform using the following equation:

\[
F^{-1}[f(u, v)] = I(r, c) = \frac{1}{N} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u, v) e^{j \frac{2 \pi (ur + vc)}{N}}
\]

(2-10)

2.7.1 Fourier Analysis

Fourier's representation of functions as a superposition of sines and cosines has become ubiquitous for both the analytic and numerical solution of differential equations and for the analysis and treatment of communication signals. Fourier and wavelet analyses have some very strong links.
2.7.2 Discrete Fourier Transform

The discrete Fourier transform (DFT) estimates the Fourier transform of a function from a finite number of its sampled points. The sampled points are supposed to be typical of what the signal looks like at all other times.

The DFT has symmetry properties almost exactly the same as the continuous Fourier transforms. In addition, the formula for the inverse discrete Fourier transform is easily calculated using the one for the discrete Fourier transform because the two formulas are almost identical.

2.7.3 Windowed Fourier Transforms

If $f(t)$ is a non periodic signal, the summation of the periodic functions, sine and cosine, does not accurately represent the signal. You could artificially extend the signal to make it periodic but it would require additional continuity at the endpoints. The windowed Fourier transform (WFT) is one solution to the problem of better representing the non-periodic signal. The WFT can be used to give information about signals simultaneously in the time domain and in the frequency domain.

With the WFT, the input signal $f(t)$ is chopped up into sections, and each section is analyzed for its frequency content separately. If the signal has sharp transitions, we window the input data so that the sections converge to zero at the endpoints [28]. This windowing is accomplished via a weight function that places less emphasis near the interval's endpoints than in the middle. The effect of the window is to localize the signal in time.
2.7.4 FAST FOURIER TRANSFORMS

To approximate a function by samples, and to approximate the Fourier integral by the discrete Fourier transform, requires applying a matrix whose order is the number sample points $n$: Since multiplying an $n \times n$ matrix by a vector costs on the order of $n^2$ arithmetic operations, the problem gets quickly worse as the number of sample points increases. However, if the samples are uniformly spaced, then the Fourier matrix can be factored into a product of just a few sparse matrices, and the resulting factors can be applied to a vector in a total of order $n \log n$ arithmetic operations. This is the so-called fast Fourier transform or (FFT), [29].
2.8 Similarities between Fourier and Wavelet Transforms

The fast Fourier transform (FFT) and the discrete wavelet transform (DWT) are both linear operations that generate a data structure that contains log 2 n segments of various lengths, usually filling and transforming it into a different data vector of length 2 n.

The mathematical properties of the matrices involved in the transforms are similar as well. The inverse transform matrix for both the FFT and the DWT is the transpose of the original. As a result, both transforms can be viewed as a rotation in function space to a different domain. For the FFT, this new domain contains basis functions that are sines and cosines. For the wavelet transform, this new domain contains more complicated basis functions called wavelets, mother wavelets, or analyzing wavelets.

Both transforms have another similarity. The basis functions are localized in frequency, making mathematical tools such as power spectra (how much power is contained in a frequency interval) and scalegrams useful at picking out frequencies and calculating power distributions.

2.9 Dissimilarities between Fourier and Wavelet Transforms

The most interesting dissimilarity between these two kinds of transforms is that individual wavelet functions are localized in space. Fourier sine and cosine functions are not. This localization feature, along with wavelets' localization of frequency, makes many functions and operators using wavelets ‘sparse’ when transformed into the wavelet domain. This sparseness, in turn, results in a number of useful applications such as data compression, detecting features in images, and removing noise from time series.

One way to see the time-frequency resolution differences between the Fourier transform and the wavelet transform is to look at the basis
function coverage of the time-frequency plane [30]. Figure (2-1) shows a windowed Fourier transform, where the window is simply a square wave. The square wave window truncates the sine or cosine function to fit a window of a particular width. Because a single window is used for all frequencies in the WFT, the resolution of the analysis is the same at all locations in the time-frequency plane.

Figure (2-1). Windowed Fourier transform.

An advantage of wavelet transforms is that the windows vary. In order to isolate signal discontinuities, one would like to have some very short basis functions. At the same time, in order to obtain detailed frequency analysis, one would like to have some very long basis functions. A way to achieve this is to have short high-frequency basis functions and long low-frequency ones. This happy medium is exactly what you get with wavelet transforms. Figure (2-2) shows the coverage in the time-frequency plane with one wavelet function, the Daubechies wavelet.

One thing to remember is that wavelet transforms do not have a single set of basis functions like the Fourier transform, which utilizes just
the sine and cosine functions. Instead, wavelet transforms have an infinite set of possible basis functions. Thus wavelet analysis provides immediate access to information that can be obscured by other time-frequency methods such as Fourier analysis.

Figure (2-5). Daubechies wavelet basis functions, time-frequency tiles, and coverage of the time-frequency plane
CHAPTER THREE

IMAGE CLASSIFICATION
Chapter Three
Image Classification

3.1 Introduction

The classification of satellite image is a prime component of any quantitative analysis process. The process of classification involves labeling of each pixel of image into a ground cover type using its numeric values in different spectral channels. In other words, image classification is the process of creating thematic maps from satellite imagery, a thematic map is an informational representation of an image which shows the spatial distribution of a particular class. Classes can be as diversified as their areas of interest for example soil, vegetation, water, and clouds. Inside a class, can be defined as subclasses, and thus the process of classification needs to become more refined [6].

The mathematical technique of pattern recognition is used to classify each pixel of the image corresponding to specific ground cover types. Here the pattern means a pixel itself or a vector, which has elements from all available spectral bands in the form of brightness values arranged in column form. Supervised classification and unsupervised classification are the two broad types of classification processes that are used in satellite remote sensing.

3.2 Simple Pixel Based Classifiers

Simple pixels based classifiers originated in the 1970’s, and were designed for multispectral data. They can be divided into two different types, Supervised and Unsupervised. Supervised classifiers require the user to decide which classes exist in the image, and then to delineate samples of these classes. These samples (known as training areas) are then input into a classification program, which produces a classified image. Unsupervised classification does not require training areas, just
the number of classes you would like to end up with. You should be aware, though, that the classes an unsupervised classifier creates may be quite different from the classes a human would identify [34].

3.3 Supervised Classification

Supervised classification is the most important technique used for the extraction of quantitative information from a satellite image. Supervised classification is much more effectual in terms of accuracy in mapping considerable classes whose validity depends largely on the cognition and skills of the image specialist. The technique assumes that each spectral class in the image can be described by a probability function in multi-spectral space [6].

Some of the supervised techniques does not uses probability distribution and use some other kind of mathematical discriminate functions. Maximum Likelihood Classification, Minimum Distance Classification, Parallelepiped Classification and Mahalanobis Classifications come under supervised classification techniques.

3.3.1 Minimum-Distance-to-Means

This type of classifier determines the mean value (DN) of each class in each band. It the assigns unknown pixels to class whose means are most similar to the value of the unknown pixel (Fig 3-1). This method is quite efficient, which made it good choice before the advent of modern computers. It may still be worth considering if you have a very large image.
3.3.2 Parallelepiped Classifier

This classifier is a bit more sophisticated than the minimum-distance-to means classifier. It works by delineating the boundaries of a training class using straight lines. This can most easily be visualized in the simple case where only two bands are used. In this case the boundaries look like a series of rectangles Figure (3-2). After the boundaries have been set, unknown pixels are assigned to a class if they fall within the class's boundaries. If the unknown pixel does not fall within any class's boundary, it is classified as unknown. This method is computationally efficient, and attempts to capture the boundaries of each class. However, using straight lines to delineate the classes limits the method's effectiveness. Also, having pixels classified as unknown may be undesirable for some applications.
3.3.3 Maximum Likelihood

Maximum likelihood classification is by far the most popular type of classification for multispectral data. It first determines the distributions of the DN values in each band for each class. Each unknown pixel is then assigned to a class based upon Gaussian probability (Fig 3-3). This method should produce better results than the previous two methods. Although it is an expensive, modern computers allow it to be widely used.

3.4 Unsupervised Classification

In this technique pixels in an image are assigned to spectral classes without the user having fore knowledge of the existence or names of those classes. This technique is performed using clustering methods.
Clustering can be used to identify number of classes in which pixels of image can fall. This method is time consuming. This technique is useful for determining the spectral class composition of the data prior to detailed analysis by the methods of supervised classification [6].

3.4.1 K-Means
   This method works by choosing random seeds, which can be though of as points with random DN values. After the seeds have been chosen lines are formed to separate the classes. Next, the points lying within the delineated areas are analyzed, and their means are noted. The means then form the new seeds, and a new series of lines are formed to separate the classes. This process is then repeated several times.

3.4.2 Fuzzy C Means
   Very similar to K-Means, but fuzzy logic is incorporated.

3.4.3 ISODATA
   A more sophisticated version of the K-Means classifier which allows classes to be created and destroyed.

3.5 Hyperspectral Classifiers
   One of the problems facing people using hyperspectral data is that the traditional multispectral classifiers cannot be used. This is because the accuracy of the supervised classifiers discussed above decreases when you add highly correlated bands. The result has been the development of a new set of classifiers.

3.5.1 Band Selection
   The simplest way to classify an image is to use one of the multispectral classification techniques, but to use only a few of the hyperspectral bands. Choosing the bands can be done in one of two ways. Possibly the better way is to choose the bands based upon knowledge of
the spectral signatures of the objects you wish to classify. More commonly, though, one of the many algorithms available for choosing the bands that best separate the classes is used.

3.5.2 Spectral Unmixing

This method attempts to determine the fraction of each class present in the pixels. To use this method, you must know the spectral signature of each class. Spectral signatures can be obtained from either laboratory or field measurements, or may come directly from the image. A method for determining which classes are present in an unknown image will be discussed. One variant of this technique you should be aware of is Multiple Endmember Spectral Mixture Analysis (MESMA), which uses an expert system to assist the unmixing. It is useful when many classes are present.

3.5.3 Spectral Angle Mapping (SAM)

It is easiest to imagine this method in the simple case where only two bands are used. Assign one band to each axis of a scatter plot, and then plot the DN values from the training areas. In most cases you will find that you will be able to draw a line through the points from each training area to the axis (Fig 3-4). SAM works by drawing the lines for all classes, and then computing the angle of the line relative to the x-axis. A pixel is assigned to a class by plotting it, and then drawing a line through the point to the axis. Its angle relative to the x-axis can then be determined, and it is assigned to the class with the most similar angle. Although it is difficult to visualize, most applications of SAM extend this approach from two dimensions to tens or hundreds of dimensions, [35].
3.6 Texture Classification using DWB.

As mentioned above, a wavelet packet decomposition capable of providing maximum inter-class discrimination power would be the most suitable representation for a given image in a texture analysis framework. However, it cannot always be guaranteed that using more subbands directly translates to smaller classification error. Experiments demonstrate that using only a few of the subbands instead of all of the wavelet subbands can result in smaller error rates, where error rate is the ratio of total number of misclassifications and total number of pixels with the ratio expressed as a percentage. The subbands were chosen by heuristically selection, whereby subbands with apparent difference in magnitudes of the transform coefficients for different texture regions are given priority over those which do not react very strongly to one texture or other.

![Figure (3-4) Spectral Angle Mapping (SAM)](image)
There are \( n^c_k = \binom{n}{k} \) possible combinations of \( k \)  \hspace{1cm} (3-1)

Subbands from a total of \( n \) subbands. It is not practical to employ a brute force approach, which finds out the best combination by trying out each of them. This is motivation enough for finding out an efficient way of determining which of these combinations of subbands is optimal in terms of best discriminating different textures. Reduction in dimensionality of the problem may result in not only more accurate but also faster classification.

3.6.1 Discriminate Measure

Consider a wavelet packet subband node \( \lambda_{d}^{p,q} \), where \( d \) is the depth and \( p, q \) represents the location at depth \( d \) of the wavelet packet tree. We use the convention that in case of an image, a subband \( \lambda_{d}^{p,q} \) is decomposed into four subbands \( \lambda_{d+1}^{2p,2q} \), \( \lambda_{d+1}^{2p+1,2q} \), \( \lambda_{d+1}^{2p,2q+1} \), \( \lambda_{d+1}^{2p+1,2q+1} \) and \( \lambda_{0,0} \) denotes the root node (original image). Let \( f_{d}^{p,q} \) and \( g_{d}^{p,q} \) denote the normalized energy distributions of wavelet packet coefficients corresponding to the subband node \( n_{d}^{p,q} \) associated with classes 1 and 2 respectively given by[36,37];

\[
 f_{d}^{p,q}(x, y) = \frac{(w_{d}^{p,q}(x, y)c^{(1)})^2}{\|c^{(1)}\|^2} \hspace{1cm} (3-2)
\]
\[
 g_{d}^{p,q}(x, y) = \frac{(w_{d}^{p,q}(x, y)c^{(2)})^2}{\|c^{(2)}\|^2} \hspace{1cm} (3-3)
\]

Where \( w_{d}^{p,q}(x, y) \) denotes the basis vector corresponding to position \((x, y)\) in the subband \( n_{d}^{p,q} \) and \( c^{(1)} \) and \( c^{(2)} \) denote texture images corresponding to classes 1 and 2 respectively. A discriminate measure
\( D_d^{p,q}(f, g) \) should be able to measure how differently \( f \) and \( g \) are distributed thus relating it directly to the discrimination power of subband \( n_d^{p,q} \). The Kullback-Leibler distance, also known as the relative entropy, between \( f \) and \( g \) is given by:

\[
I_d^{p,q}(f, g) = \sum_x \sum_y f(x, y) \log \frac{f(x, y)}{g(x, y)}.
\]

A symmetric version of this distance measure, also known as the \( J \)-divergence, given by:

\[
D_d^{p,q}(f, g) = I_d^{p,q}(f, g) + I_d^{p,q}(g, f).
\]

is proposed to measure the discrimination power of a subband.
CHAPTER FOUR
RESULTS ANALYSIS
Chapter Four  
Results Analysis

4.1 Introduction
In this stage of the work, the classification algorithm has been tested using written program. The classification results were compared with other well-known standard classification methods, using the remote sensing and image-processing package "ENVI version 3.2". The following steps can summarize the testing operation.

4.2 Adaptive Method Results
A single band of 256×256-size satellite image is classified using the discrete wavelet transform and our adopted "Kullback-Leibler" distance algorithm, as illustrated by the flowchart:

```
Input Image
Enter name of wavelet transform levels
Selections the number of regions of interest (ROI)
Select the block size
Select the number of bands to used
Show the correlation factors between each band and others
Compute the Kullback-Leiber distances between the selected image bands
Out put images classified images
```
1- An image of standard format "JPEG or/and BMP" is first entered.

2- The name of the wavelet transform file and the desired number of decomposition levels of the wavelet transform (i.e. 1, 2, 3 etc) should be identified in the appeared input message boxes. The results are (4, 7, 10) images, respectively, each has the same sizes as the original input image.

3- Selection of the number of the regions of interest "ROI" is performed easily by utilizing the mouse clicking on the menu "Selection" first, and chosen the desired number of regions from the appeared message box then.

4- The window's size used to occupy the desired elements of each ROI, and then for the matching process is selected by the appeared message box. Only odd sizes are allowable to be certain that clicked pixel is adopted with the uniform surrounded elements.

5- The wavelet transformation, as described above, produces certain number of image bands corresponding to the chosen transformation level. For classification purposes, not necessarily all the produced bands are utilizes.

6- As all the above mentioned parameters were defined, the program is designed to determine, automatically, the correlation factors between each band and others. For example, when three bands were adopted, fifteen correlation parameters should be
computed and displayed in the list box, discussed in chapter three.

7- The next step is to performing the computation of Kullback-Leibler distances (given in section 3.2) between the selected image bands. As it is shown, if the selected number of bands were three and the selected number of regions of interest were five, the output correlation distances will be fifteen.

8- As it is shown in Table (4-1), number of points assigned for each class can be counted to be compared with other classification method.

Different transformation levels and block sizes have been tried in our present research, denoted as C1, to C9 below:

C1: wavelet level =1, BS (block size) = 3
C2: wavelet level =1, BS (block size) = 5
C3: wavelet level =1, BS (block size) = 7
C4: wavelet level =2, BS (block size) = 3
C5: wavelet level =2, BS (block size) = 5
C6: wavelet level =2, BS (block size) = 7
C7: wavelet level =3, BS (block size) = 3
C8: wavelet level =3, BS (block size) = 5
C9: wavelet level =3, BS (block size) = 7
Table (4-1): representing the classified number of pixels for each region, using different block's sizes

<table>
<thead>
<tr>
<th>Case number</th>
<th>R1</th>
<th>R2</th>
<th>R3</th>
<th>R4</th>
<th>R5</th>
<th>R6</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>2138</td>
<td>4927</td>
<td>14989</td>
<td>9570</td>
<td>20454</td>
<td>12438</td>
</tr>
<tr>
<td>C2</td>
<td>5224</td>
<td>1890</td>
<td>14462</td>
<td>7067</td>
<td>18117</td>
<td>16744</td>
</tr>
<tr>
<td>C3</td>
<td>6476</td>
<td>419</td>
<td>14241</td>
<td>5646</td>
<td>17952</td>
<td>17766</td>
</tr>
<tr>
<td>C4</td>
<td>2133</td>
<td>4951</td>
<td>13072</td>
<td>12658</td>
<td>14431</td>
<td>17271</td>
</tr>
<tr>
<td>C5</td>
<td>5028</td>
<td>2086</td>
<td>12892</td>
<td>10786</td>
<td>12249</td>
<td>20463</td>
</tr>
<tr>
<td>C6</td>
<td>6458</td>
<td>433</td>
<td>13609</td>
<td>8096</td>
<td>13628</td>
<td>20276</td>
</tr>
<tr>
<td>C7</td>
<td>2091</td>
<td>4948</td>
<td>12065</td>
<td>13040</td>
<td>7618</td>
<td>24754</td>
</tr>
<tr>
<td>C8</td>
<td>4920</td>
<td>2164</td>
<td>12325</td>
<td>10921</td>
<td>9142</td>
<td>24032</td>
</tr>
<tr>
<td>C9</td>
<td>6399</td>
<td>510</td>
<td>13078</td>
<td>8760</td>
<td>10540</td>
<td>23213</td>
</tr>
</tbody>
</table>

The classification results are demonstrated in figure (4-1 C1 to C9) below.
Figure (4-1): Demonstration of the classification results with different adopted bands (i.e. transformation levels) and block's sizes

Figure (4-2): Illustrating the relationship between number of classified points within each class, using different block's sizes (i.e. C1=3, C2=5, and C3=7), for wavelet transform of one-level.
Figure (4-3): Illustrating the relationship between classified points within each class with different levels of wavelet transformation (i.e. C1= level-one, C4= level-two, and C7=level-three), using fixed block size=3.

Figure (4-4): Illustrating the relationship between classified points within each class with different levels of wavelet transformation (i.e. C2= level-one, C5= level-two, and C8=level-three), using fixed block size=5.
4.3 Other classification methods results

To compare our obtained results with other well known supervised classification algorithms, the ENVI 3.2 package has been utilized to perform the classification procedures, using the Minimum Distance and Parallelepiped algorithms. Regions in each classified image are shown as listed below.

Region 1    Water       blue
Region 2    Rock        cyan
Region 3    Veg         green
Region 4    Road        red
Region 5    Urban       yellow
Region 6    Soil        brown
Region 0    unsupervised black

Figure (4-13) represents the classified image by performing the supervised minimum distance classifier. Number of points assigned to each class is listed in Table (4-2).
Figure (4-6): Classification results by minimum distance classifier, using single-band image

Table (4-2): represents the number of points and the decided color for each class

<table>
<thead>
<tr>
<th>Class #</th>
<th>No. of points</th>
<th>Class color</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8598</td>
<td>blue</td>
</tr>
<tr>
<td>2</td>
<td>2970</td>
<td>cyan</td>
</tr>
<tr>
<td>3</td>
<td>11211</td>
<td>green</td>
</tr>
<tr>
<td>4</td>
<td>12472</td>
<td>red</td>
</tr>
<tr>
<td>5</td>
<td>4076</td>
<td>yellow</td>
</tr>
<tr>
<td>6</td>
<td>26209</td>
<td>brown</td>
</tr>
</tbody>
</table>

The supervised Parallelepiped classification result is illustrated in figure (4-15). Number of points assigned for each classified region is given in Table (4-3).
Figure (4-7): Classification result obtained by performing the supervised parallelepiped algorithm, using single-band image

Table (4-3): represents the number of points and the decided color for each class

<table>
<thead>
<tr>
<th>Class #</th>
<th>No. of points</th>
<th>Class color</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7812</td>
<td>blue</td>
</tr>
<tr>
<td>2</td>
<td>30941</td>
<td>cyan</td>
</tr>
<tr>
<td>3</td>
<td>22382</td>
<td>green</td>
</tr>
<tr>
<td>4</td>
<td>1964</td>
<td>red</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>yellow</td>
</tr>
<tr>
<td>6</td>
<td>2437</td>
<td>brown</td>
</tr>
</tbody>
</table>

Table (4-3)
The same classification algorithms (i.e. minimum distance and parallelepiped classifiers) have been implemented on six image bands. The results are demonstrated in figures (17 to 20) and tables (4-4 and 4-5) below.

Figure (4-8): Minimum distance classification result, using 6-bands

Table (4-4): represents the number of points and the decided color for each class, using the minimum distance classifier on 6-bands

<table>
<thead>
<tr>
<th>Class #</th>
<th>No. of points</th>
<th>Class color</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7801</td>
<td>Blue</td>
</tr>
<tr>
<td>2</td>
<td>1407</td>
<td>Cyan</td>
</tr>
<tr>
<td>3</td>
<td>3461</td>
<td>Green</td>
</tr>
<tr>
<td>4</td>
<td>20558</td>
<td>Red</td>
</tr>
<tr>
<td>5</td>
<td>4843</td>
<td>Yellow</td>
</tr>
<tr>
<td>6</td>
<td>27461</td>
<td>Brown</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>Black</td>
</tr>
</tbody>
</table>
Table (4-5): represents the number of points and the decided color for each class, using the parallelepiped classifier on 6-bands

<table>
<thead>
<tr>
<th>Class #</th>
<th>No. of points</th>
<th>Class color</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6875</td>
<td>Blue</td>
</tr>
<tr>
<td>2</td>
<td>30651</td>
<td>Cyan</td>
</tr>
<tr>
<td>3</td>
<td>20094</td>
<td>Green</td>
</tr>
<tr>
<td>4</td>
<td>4687</td>
<td>Red</td>
</tr>
<tr>
<td>5</td>
<td>627</td>
<td>Yellow</td>
</tr>
<tr>
<td>6</td>
<td>2360</td>
<td>Brown</td>
</tr>
<tr>
<td>0</td>
<td>242</td>
<td>Black</td>
</tr>
</tbody>
</table>
4.4 Comparison between the Methods

In another step the classification results using the new technique were compared with the standard classification methods thorough the well known remote sensing and image processing program ENVI version 3.2. The comparison results were illustrated in the following figures. The tested case of our new technique was C3 because it represented the best classification results due to sharp boundary between classes and the small unclassified regions.

Figures (4-21 to 24) represents comparisons between our classification method with the minimum distance and parallelepiped classifiers, using single and multi-bands images.

Figure (4-10): Comparison between C3 (Table 4-1), and the minimum distance classifier (one-level wavelet transformation, and block size=7).
Figure (4-11): Comparison between C3 (Table 4-1), and the parallelepiped classifier (one-level wavelet transformation, and block size=7).

Figure (4-12): Comparison between C3 (Table 4-1), and the minimum distance classifier (one-level wavelet transformation, and block size=7), implemented on 6-band images.
Figure (4-13): Comparison between C3 (figure 4-7), and the parallelepiped classifier (one-level wavelet transformation, and block size=7), implemented on 6-band images..
CHAPTER FIVE
CONCLUSIONS AND SUGGESTIONS
FOR FUTURE WORKS
CHAPTER FIVE
CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORKS

5.1 Conclusion

- As it has been seen in previous chapter, increasing of block size yield better classification results due to the enhancement of the clustering or discrimination power which, in turns, avoiding the possibility in the interference of the classes. This problem was obvious when the selected block size was $3 \times 3$.

- As the Wavelet transform level increases, the classification results remain approximately the same even if the number of bands increased, when the adopted block size was unchanged.

The points concluded above have been illustrated well in chapter four, figures [(4-2) – (4-5)]; as flowering;

1. Fig (4-2), cases C1, C2, and C3 showed increasing block size for the same transformed level has been improved as the adopted block size was increased.

2. Fig (4-3), for cases C1, C4, C7, showed that increasing the transformation levels did not affected as the block size remain unchanged (block size = $3 \times 3$).

3. Fig (4-4), for cases C2, C5, C8, showed that increasing the transformation levels did not affected as the block size remain unchanged, for (block size = $5 \times 5$).
4. Fig (4-5), for cases C3, C6, C9, showed that increasing the transformation levels did not affect as the block size remain unchanged, for (block size=7×7).

- From the points discussed above, we can say that the situation illustrated in figure (4-1) yielded the best classification results. This result was found as to be very close to the Minimum Distance supervised classification technique implemented on multi-band images (i.e. 6-bands). However, the conclusion can be demonstrated with the help of the illustrated of figures [(4-10) – (4-13)], and as follows:

1. Fig (4-10) showed the comparison between C3 & the minimum distance classifier.

2. Fig (4-11) showed the comparison between C3 and the parallelepiped classifier.

3. Fig (4-12) showed the comparison between C3 and the minimum distance classifier, implemented on 6-band images.

4. Fig (4-13) showed the comparison between C3 and the parallelepiped classifier, implemented on 6-band images.

- This method is very usefully to classify the single band image such as, high resolution satellite image, weathering satellite images.
5.2 Suggestions for Future Works

Despite the good segmented classified regions obtained by our introduced method, may new results can be achieved if the following suggestions may be adopted;

1. The K-L classification can be use with Principal component analysis to classify the enhance PC or other PCs.

2. Minimum-Noise-Filtering MNF may be used to reduce the noise effects from image bands before performing the wavelet and classification processes.

3. The classification accuracy can be improved by modify the software to evaluate results in multi iteration.
REFERENCES


19-"SAR image classification by wavelet-based texture analysis" Punya Thitimajshima, 1999.


الخلاصة

تستخدم طرق تصنيف مختلفة لعمليات تحليل صور الأقمار الصناعية المتعددة الحزم، وتشير تلك المشاكل عند استخدام صورة مفردة الحزمة (single band image). في هذا البحث التحليل الموبيجي الساكن (stationary-wavelet transforms) يقوم بإنتاج صور متعددة الحزم من الصور الأحادية الحزمة المستخدمة (بحيث تكون الصور المتولدة لدينا بنفس حجم الصورة الأصلية).

من المعروف أن طرق التصنيف التي تستعمل عادة في التحسس النائي تقسم إلى مرشدة (supervised) وغير مرشدة (unsupervised). في الأغلب تقنيات التصنيف أما تعتمد على التعقد (clustering) ، أو على الإحصائيات (statistical-features).

بالنسبة للعلماء المختصين بالمعالجة الصورية فمهم المعروف بأن الأنترولي (entropy) يشير إلى كمية المعلومات الموجودة في الصورة. في بحثنا المقدم نعتمد على الأنترولي النسبي للصورة المفردة الحزمة (relative entropy) التي تم تحويلها إلى متعددة الحزم بواسطة التحويل لتطبيق التصنيف. التحويل المستخدم في هذا البحث هو التحويل الموبيجي الساكن (stationary-wavelet) الذي ينتج صور متعددة متساوية في الحجم مع الصورة الأصلية.

وتقدم كبديل Kullback-Liebler تسمى هذه الطريقة أيضاً لاستخدامها في التحسس النائي. الطريقة تنفذ على الصورة المفردة الحزمة المتعددة الحزم باستخدام حجم خطوة (block size) مختلف.

وتغير عدد التكرارات للتحويل الموبيجي وحجم الخطوة نحصل على نتائج تتم مقارنتها مع تقنيات التصنيف المتبعة عادة (i.e. Minimum distance and parallelepiped). وقد أثبتت النتائج بأن طريقتنا المقدمة تقيمتها تعطي نتائج طريقة لطريقة minimum distance. طريقتنا المقدمة انجريت بواسطة برنامج Visual-basic بعدد من الخطوات الروتينية لإنتاج التصنيف. مناطق عريضة تمثل مناطق الصورة تم
إفرازها بواسطة طريقتنا المقدمة تقارن مع النتائج المستحصلة من تقنيات التصنيف الأخرى (باستخدام البرمجيات الجاهزة مثل ENVI).