Advanced Methods in Remote Sensing

Lectures 9-12: Wavelet Analysis

Part 1: Background to Integration and Convolution

Recommended Reading:

Advanced Engineering Mathematics, Erwin Kreyszig (5th + Edition) Wiley.
→ A very good (and not too advanced) general mathematics textbook

Summary

This lecture will introduce you to the concept of wavelet analysis and how it might be useful to remote sensing and ecological problems

Essential Mathematics

Wavelet analysis relies on a solid foundation in mathematics. Therefore we will take the time to refresh/introduce you to:

- Integration
- Convolution

If you are interested in pursuing this subject past that described in this lecture – I can point you in the direction of books and papers.

Remember: Mathematics is all about rules and tricks.

Integration

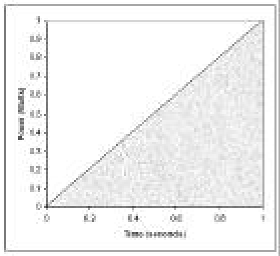
What it does:

Calculates the area underneath a curve: e.g. in a graph of drill power with time the area under the curve = energy used

The Basic RULE of Integration:

$X^n = (1/n+1)x^{n+1}+C$

Lets do this example:



The Rules of Integration

$k, cons$	$kx + c$
x^n	$\frac{x^{n+1}}{n+1} + c \quad (n \neq -1)$
$x^{-1} = \frac{1}{x}$	$\ln\lvert x \rvert + c$
e^{ax}	$\frac{e^{ax}}{a} + c$
$\sin(ax+b)$	$\frac{-\cos(ax+b)}{a} + c$
$\cos(ax+b)$	$\frac{\sin(ax+b)}{a} + c$

$$\int u\left(\frac{dv}{dx}\right)dx = uv - \int v\left(\frac{du}{dx}\right)dx$$

Convolution

Convolution is central to using wavelet analysis and can be mathematically expressed as the following integral:

$$g(x) = f(x) \otimes h(x) = \int_{-\infty}^{\infty} f(s)h(x-s)\,ds$$

But WHAT DOES THIS MEAN!!!

Convolution measures the AREA of overlap between one function, f(x) and the spatially reversed (I.e. mirror image) version of the other function, g(x).

Effectively: How similar are two functions over all spatial locations

To work it out you multiply each value of the operator with that of the signal and then add up all these values.

What About Correlation?

Correlation is similar to convolution EXCEPT you do not 'flip or reverse' the function in space:

$$g(x) = f(x) \otimes h(x) = \int_{-\infty}^{\infty} f(s)h-(s-x)ds$$

Where h-(x) is the complex conjugate of h(x).

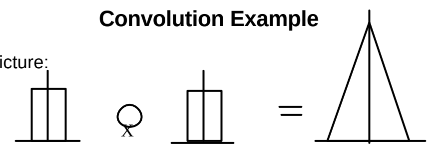
But WHAT DOES THIS MEAN!!!

Correlation again measures the AREA of overlap between one function, f(x) and another function, g(x).

Effectively: A Measure of the direct (untransformed) similarity of the two functions.

Convolution Example

Picture:

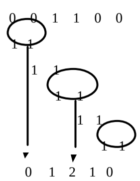


Math:

i.e. 1*0 + 1*0 = 0

i.e. 1*1 + 1*1 = 2

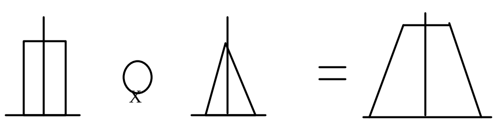
i.e. 1*0 + 1*0 = 0



i.e. the convolution of two top-hat functions is a big spiky triangle.

Convolution Example 2

Picture:



i.e. the convolution of a top-hat (0110) function and a triangle (0 ½ 1 ½ 0) is a broadened and slightly larger flattened triangle (0 1 1½ 1½ 1 0).

In general:

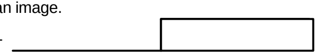
Convolution of two identical objects = big spike;

Convolution of two non-identical objects = more flattened object

Edge Detection

In image processing you frequently use convolution when passing an operator over an image.

Operator = 1 -1



0 0 0 0 1 1 1 1

First Flip the Operator: i.e. -1 1 - Then multiply each element and add them together:

0 0 0 0 1 1 1 1

i.e. -1*0 + 1*0 = 0 -1 1

-1 1

-1 1

i.e. -1*0 + 1*1 = 1 -1 1

-1 1

i.e. -1*0 + 1*0 = 0 -1 1

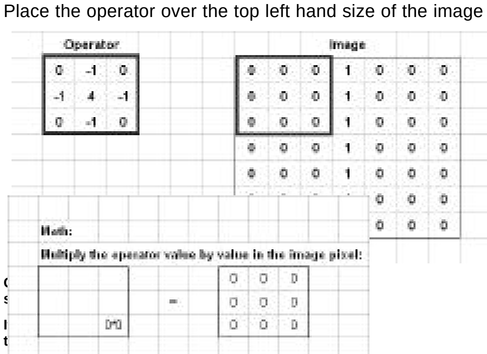
0 0 0 1 0 0 0

An Edge becomes a Spike

This is the 1st derivative

In 2D: Step 1

Place the operator over the top left hand size of the image



Operator

0	-1	0
-1	4	-1
0	-1	0

Image

0	0	0	1	0	0	0
0	0	0	1	0	0	0
0	0	0	1	0	0	0
0	0	0	1	0	0	0
0	0	0	1	0	0	0
				0	0	0
				0	0	0

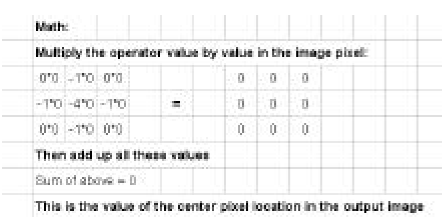
Math:

Multiply the operator value by value in the image pixel:

				0	0	0
			=	0	0	0
		0*0		0	0	0

Step 2:

Add up all the answers from the multiplications



Math:

Multiply the operator value by value in the image pixel:

0*0	-1*0	0*0		0	0	0
-1*0	-4*0	-1*0	=	0	0	0
0*0	-1*0	0*0		0	0	0

Then add up all these values

Sum of above = 0

This is the value of the center pixel location in the output image

Step 3:

Repeat this process by moving the operator over all possible locations in the image

Step 2:

Again Add up all the answers from the multiplications

Math:

Multiply the operator value by value in the image pixel:

0*0	-1*1	0*0		0	-1	0
-1*0	4*1	-1*0	=	0	4	0
0*0	-1*1	0*0		0	-1	0

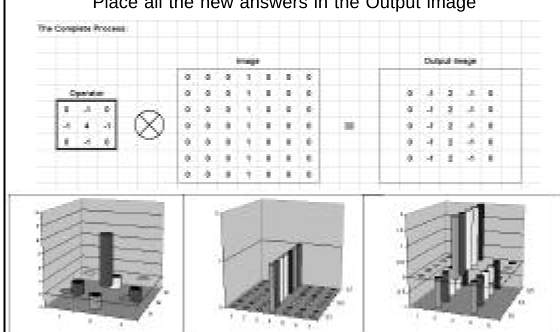
Then add up all these values

Sum of above = 2

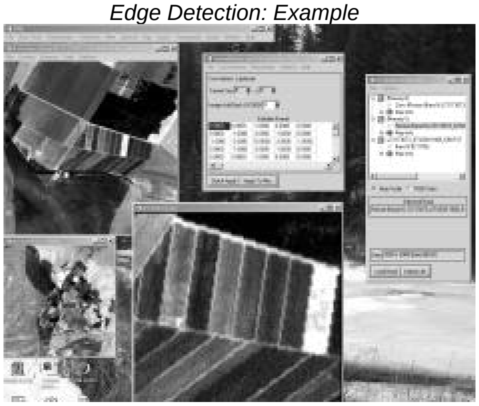
This is the value of the center pixel location in the output image

Step 4:

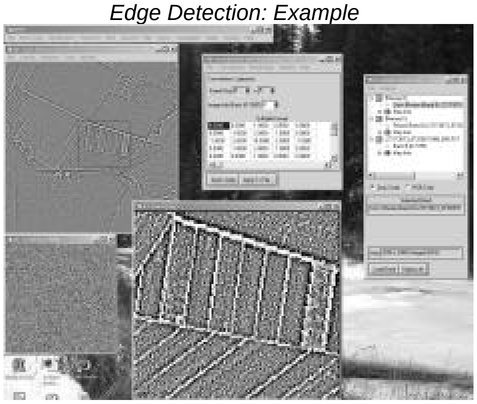
Place all the new answers in the Output image



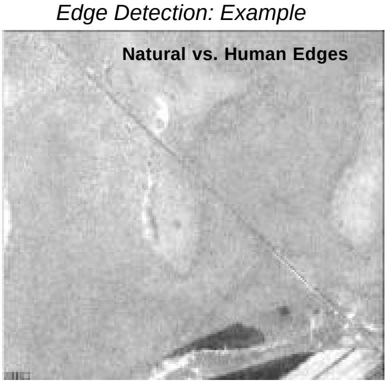
Edge Detection: Example



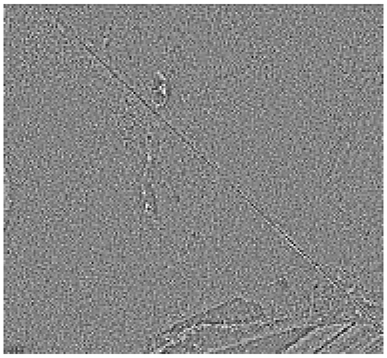
Edge Detection: Example



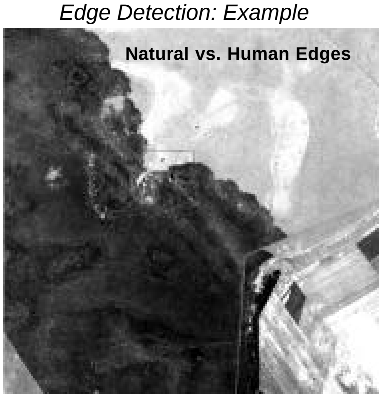
Edge Detection: Example



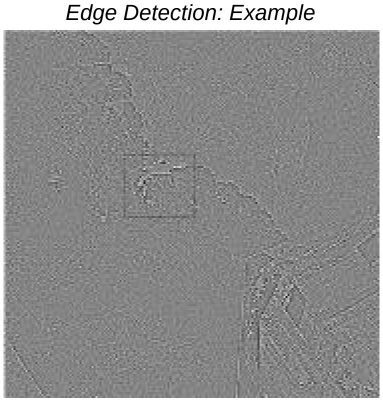
Edge Detection: Example



Edge Detection: Example



Edge Detection: Example



Advanced Methods in Remote Sensing

Lectures 9-12: Wavelet Analysis

Part 2: Background to Fourier Analysis

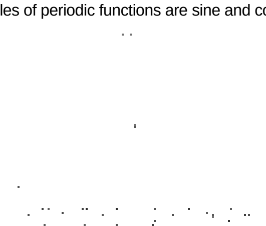
Periodic Functions

A function f(x) is periodic over all values of x, if for a given constant C:

$$f(x + C) = f(x)$$

C is called the ***period*** of f(x).

Common Examples of periodic functions are sine and cosine waves:



Both cos nx and sin nx, where n = 2p are periodic functions.

The Basics of Fourier Synthesis/Analysis

Mathematical discipline began is 1807 with Joseph Fourier.

The just of Fourier Synthesis is that any 2p periodic based function can be broken up into a set of sine (or cosine) waves of varying frequency.

i.e. in Math: any 2p periodic based function can be expressed in the form:

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos\frac{2pnx}{X} + b_n \cos\frac{2pnx}{X})$$

Where:

$$a_0 = \frac{2}{X}\int_0^X f(x)dx$$

$$a_n = \frac{2}{X}\int_0^X f(x)\cos\frac{2pnx}{X}dx$$

$$b_n = \frac{2}{X}\int_0^X f(x)\sin\frac{2pnx}{X}dx$$

The series is called a 'Fourier Series'

The parameters are called 'Fourier Coefficients'

Frequency/Scale Analysis

Fourier Synthesis was the start of functional multi-scale mathematics – i.e. the analysis of functions, f(x), that vary in scale (size).

Fourier analysis allowed functions to be analyzed over a series of scales. This if often called 'frequency analysis' → or 'Scale Analysis'.

The Steps:

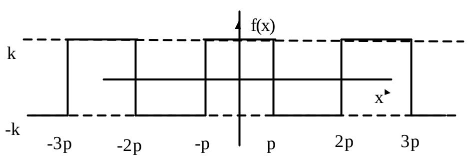
- Create a function f(x)
- Compare f(x) to another function – say g(x)
- Use this comparison to approximate the shape of g(x)
- Change the size of f(x) and compare it again to g(x).
- Repeat the procedure over a series of sizes of f(x)

Using Fourier Series to Approximate a Function

Consider the function:

$$f(x) = \begin{cases} -k & when \quad -p < x < 0 \\ k & when \quad 0 < x < p \end{cases} \quad and \quad f(x+2p) = f(x)$$

This is a SQUARE WAVE and has the following shape:



Apply Fourier Analysis to the Square Wave:

As the Area under the 'curve' is zero then $a_0 = 0$

Calculating a_n using the trigonometry relation sin nx = 0.

$$a_n = \frac{1}{p}\int_{-p}^{p} f(x)\cos nx dx =$$

$$\frac{1}{p}\left[\int_{-p}^{0}(-k)\cos nx dx + \int_{0}^{p} k \cos nx dx + \right] =$$

$$\frac{1}{p}\left(\left| -k\frac{\sin nx}{n}\right|_{-p}^{0} + \left| k\frac{\sin nx}{n}\right|_{0}^{p}\right) = 0$$

Apply Fourier Analysis to the Square Wave:

Calculating b_n using the trigonometry relations: cos (-a) = cos a and cos 0 = 1:

$$b_n = \frac{1}{p}\int_{-p}^{p} f(x)\sin nx dx =$$

$$\frac{1}{p}\left(\left|k\frac{\cos nx}{n}\right|_{-p}^{0} - \left|k\frac{\cos nx}{n}\right|_{0}^{p}\right) =$$

$$\frac{2k}{np}(1-\cos np)$$

Therefore the Fourier Coefficients b_n are:

$$b_1 = \frac{4k}{p}, \quad b_2 = 0, \quad b_3 = \frac{4k}{3p}, \quad b_4 = 0, \quad b_5 = \frac{4k}{5p}, \ldots$$

What the Approximations Look Like

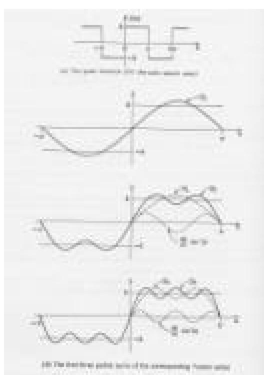
There are the plots of the 1st three (non-zero) Fourier coefficients.

The approximation is improved by adding the higher frequency components.

FROM FOURIER THEORY: Adding up ALL the coefficients in this way (i.e. not just the 1st 3) will reconstruct the original function.

This process is often called Fourier or Signal Decomposition, as the original 'Square Wave' is decomposed (split up) into a series of frequency components that each try and approximate the original 'square wave'.

Signal Decomposition is widely used in Signal and Image Processing research.



Applications of Signal Decomposition

Noise Removal:

Step 1 - Split the periodic function into each of its coefficients

Step 2 - Delete the coefficient with the highest frequency (assume =noise)

Step 3 - Add the remaining sinusoidal functions together to approximate the original function minus the noise

You could modify these steps if you wanted to isolate the noise instead of removing it.

In Summary - Fourier Analysis allows you to highlight signal features of a certain frequency (or size).

Advanced Methods in Remote Sensing

Lectures 9-12: Wavelet Analysis

Part 3: Introduction to Wavelet Analysis

From Fourier Series to Wavelet Analysis

In Fourier Series you decompose a periodic function into a linear combination of sine and cosine basis functions.

However, Why does your so called basis function have to be sinusoidal? – They don't! They just need to be orthogonal.

This is the idea behind wavelet analysis.

In 1909 Haar in his PhD thesis first hinted that discrete orthogonal basis functions or wavelets could be used to decompose a periodic function.

In the 1930s, Levy used the Haar wavelet basis to investigate Brownian Motion.

These days, wavelets are used in a wide variety of scientific fields, including medicine, astronomy, and remote sensing

Wavelet Analysis – The Basics

In wavelet analysis you use convolution to decompose your signal (or image) using discrete operators of increasing sizes called the wavelet basis functions, ? (t).

The wavelet basis must have:

1. Finite Energy

$$E = \int_{-\infty}^{\infty} | y(t) |^2 dt < \infty$$

2. The Wavelet Basis must have a mean of zero

$$\hat{y}(0) = 0$$

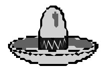
3. The FT must be real and Zero for negative frequencies

Wavelet Basis Functions

The choice of a particular wavelet basis function depends on the application. In general it should be chosen such that it approximates the shape of the objects of interest

An Example is the 2-D Mexican Hat (or Sombrero for the PC minded), which is given by:

$$\Psi(x, y) = \pm(1-x^2 - y^2) * e^{-(x2+y2)/2}$$



$\pm$ denotes whether the Hat shape is up (+) or down (-)

This original function is often called the 'mother function'. To create the set of 'wavelet basis functions' to use in the decomposition (called daughter functions) you use the following equation:

$$y_{a,b}(x,y) = \frac{1}{a} y\left(\frac{[x,y]-b}{a}\right)$$

Dilation and Translation

a is the scaling factor (and is > 0) – determines the width of the resultant wavelet daughter function.

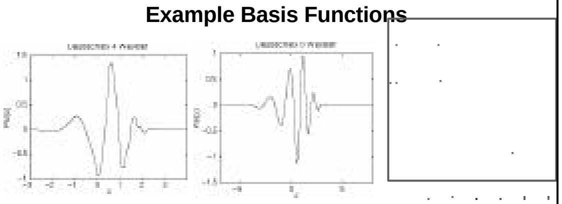
a > 1, the resultant set of daughter functions are dilated (i.e. widened) and when a < 1 the set of daughter functions are contracted

b is the translation parameter - determines the location within the image the wavelet function is centered on.

The a^{1} factor allows the energy of the wavelet to be normalized with respect to the particular scaling factor.

The role of a ensures that the shape of the function on the principal coordinates (i.e. x and y) are scaled by the same amount, so that each daughter function preserves the shape of the wavelet mother function.

Example Basis Functions



Many others exist – In general you should chose the wavelet that is closest in shape to the features you are assessing.

Wavelet Basis are Band pass Filters

When the wavelet basis meets condition 2, it also meets the criteria for a BANDPASS filter.

Result → The basis at a certain size only lets through the signal components within a certain range of frequencies (i.e. features that are of a similar size), which are defined by the energy spectrum (energy spectrum = the energy at each frequency value).

The result is that when you convolve the wavelet basis with a signal only those features of similar size to the wavelet basis are processed in the convolution.

Remember from convolution:

Convolution of two identical objects = big spike;

Convolution of two non-identical objects = more flattened object

A Question of Scale

In all forms of wavelet analysis you convolve your wavelets of increasing size with your signal or image.

Any features that are of the same size and shape to your wavelet basis function will be highlighted by a big spike.

Therefore, wavelets highlight features of a similar scale.

There exist many different shapes of wavelet basis.

Mathematically Speaking (The equivalent of 2D Analysis):

2D CWT:

$$T_C(a,b) = \int_{-\infty}^{\infty} w(x,y) y_{a,b}(x,y)\,dxdy$$

2D DWT:

$$T_{a,b} = \int_{-\infty}^{\infty} w(x,y) y_{a,b}(x,y)\,dxdy = \langle w, y_{a,b} \rangle$$

Advanced Methods in Remote Sensing

Lectures 9-12: Wavelet Analysis

Part 4: Wavelet Analysis Methods

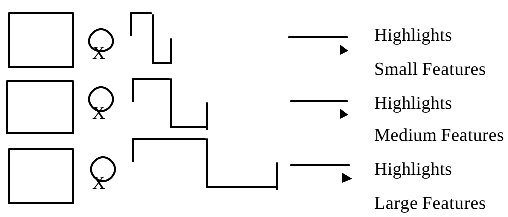
The Different Types of Wavelet Analysis

There exist several different ways to use wavelets. The most common include:

- Wavelet Decomposition
- The Wavelet Decomposition Signature – Texture Analysis
- Using the Wavelet Intensity to match feature sizes
- The Wavelet Variance Function

Wavelet Decomposition

By convolving a signal (or image) a series of wavelet bases of increasing size (and therefore decreasing frequency) you produce a wavelet decomposition:



Small Features are in general noise, while large features are broad-scale effects

Wavelet Decomposition - The Method

As the smallest wavelet size convolution (or zeroth level) highlights the high frequency features - noise, you can repeat the denoising technique we looked at in Fourier Analysis.

i.e. you can remove or isolate the signals produced by the wavelet decomposition.

Some software packages (i.e. IDL's Wavelet Toolkit)achieve this result by letting you select how many energy coefficients you want to retain.

The number of 'energy coefficients' is given by 2^N, where N is the number of convolution levels.

The Wavelet Decomposition Signature

Another way of using the wavelet decomposition is for training data in classifications.

The idea is that within a decomposition the say you have 8 levels - the first 2 levels are considered noise and the last 2 levels are considered broad image features.

Therefore, for each training image area (e.g. Forest, Grass, etc), you add up the other four levels to get an identifiable wavelet signal for each cover type.

You then run the wavelet decomposition on the rest of the image and classify each pixel by the four-level decomposition signal it is most similar to.

Several studies use this approach for Image Textural Analysis.

Wavelet Energy/Intensity

The idea behind wavelet intensities relates to the properties of convolution. If an image feature is of the same 'size' as the wavelet basis then the convolution will produce a very large value, or intensity at that point.

$$E = \int_{-\infty}^{\infty} |y(t)|^2 dt < \infty$$

Therefore, successfully convolving wavelets bases of increasing sizes with an image and noting when the largest intensity occurs will provide you with information on the actual size of features within the image.

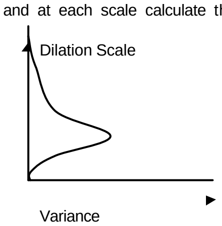
Example: Using Wavelets to Infer Object Sizes – more on this later

Wavelet Variances

Researchers have used the measure of the wavelet variance to highlight potentially important scales to analyze.

Simply run the wavelet decomposition and at each scale calculate the variance using the following equation:

$$\sigma^2 = \frac{\delta j \delta t}{C_\delta N} \sum_{n=0}^{N-1} \sum_{j=0}^{J} \frac{|W_n(s_j)|^2}{s_j}$$



The scale corresponding to the maximum variance is then assumed to contain potentially important features.

Group Exercise 1

In this exercise you as a group will consider how you could possibly use wavelet analysis to investigate Temporal Trends in Land or Sea Surface Temperatures (SST; LST)

The group is to consider:

1. What sort of data could be used to asses SST or LST?

2. What are possible factors that vary the short and long term LST?

3. How could wavelets be used to assess the periodicity of these different cycles?

4. What sort of results would you expect: i.e. are the trends hourly, daily, weekly, monthly, yearly, decadal, etc - Explain your reasoning?

After your discussion I would like you to nominate a person (who hasn't previously spoken in a class exercise to comment on each question.

Group Exercise 2

In this exercise you as a group will consider how you could possibly use wavelet analysis to determine the endmember proportions within a hyperspectral pixel

The group is to consider:

1. What mixing assumptions will you make

2. What aspect of wavelet analysis could you use? – explain your possible method

3. Would you expect wavelets to perform better than mixture modeling – if so/not why?

After your discussion I would like you to nominate a person (who hasn't previously spoken in a class exercise to comment on each question.

Advanced Methods in Remote Sensing

Lectures 9-12: Wavelet Analysis

Part 5: Ecological Applications of Wavelets

Synthesis of research conducted by E Strand, S Garrity, L Vierling, M Falkowski, AMS Smith, et al

Automatic Feature Detection Using Wavelets

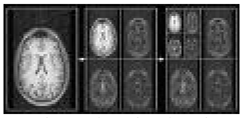
Engineering: Noise or features within signals

Image processing: Compression, reconstruction, registration

Hyperspectral data: Absorption features

Astronomy: Identify the size, shape, and location of galaxies and nebulae

Medicine: Image analysis of x-ray, MRI, and mammograms



Wavelet transformation of MRI image

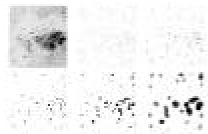
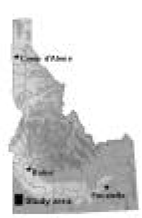


Image application in the Milky way

Can wavelets be used to automatically detect the location and crown width of trees?

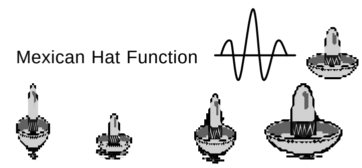
Task: Quantifying changes in juniper cover



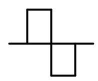
Owyhee Mountains in southwestern Idaho

The Method: Step 1

Choose a shape similar to objects of interest:

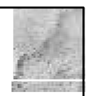


The shape must meet the following Criteria:

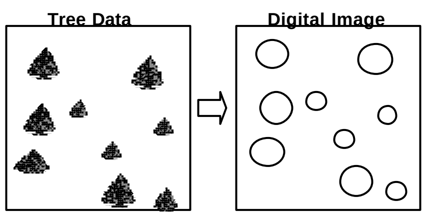


1. Area Under the Curve = 0
2. It must be finite in size

The Method: Step 2



Convert Data or use an existing Digital Image

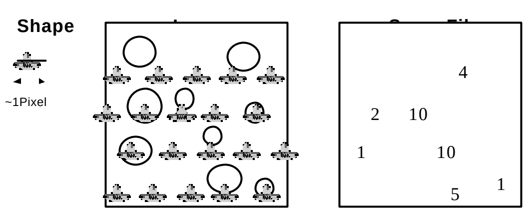


Potential Data Types include:

Aerial Photographs, Lidar, Landsat, plot data, etc

The Method: Step 3

Staring at the smallest possible Mexican Hat size: pass the Hat over the image:

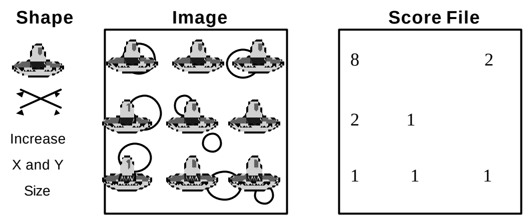


Local Maximums Only

When the shape and image object are very similar in BOTH SIZE and SHAPE a very high 'score' is recorded

The Method: Step 4

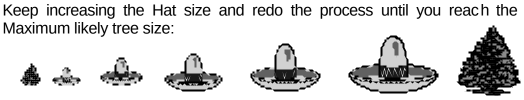
Increase the size of the Hat and repeat the process



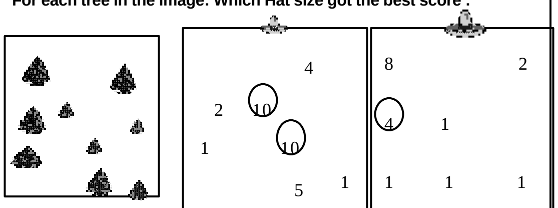
The relative ‘goodness of fit’ is recorded

The Method: Step 5

Keep increasing the Hat size and redo the process until you reach the Maximum likely tree size:



For each tree in the image: Which Hat size got the best score :



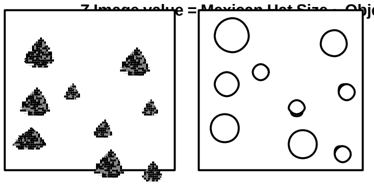
Local Maximums give tree locations. Max score gives object width

The Final Output:

Marked Point Pattern

XY center location of individual Objects

Z Image value = Mexican Hat Size – Object Width

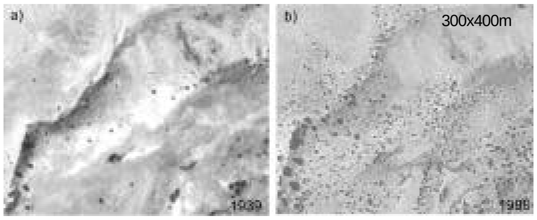


Easting	Northing	Diameter
450656	4765903	5.4
450890	4765105	8.1
450259	4766234	2.3
451360	4766790	10.2
451567	4766993	4.4
...	...	...

Output table in MATLAB

You can use a GIS to project the tree widths as circles around each point.

Application of method to aerial photograph of western juniper:



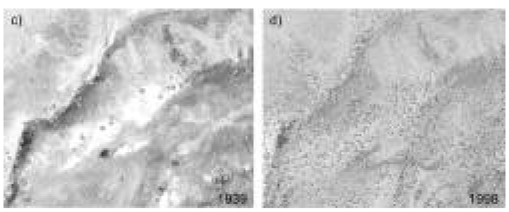
Owyhee Mountains (*Artemisia arbuscula*/*Juniperus occidentalis* habitat type)

Dark Junipers on a Sagebrush Landscape

Image Resolution = 1m

Strand *et al.* (in press)

Application: Change in Cover From 1939 to 1998



Cover 1939: 2.7%

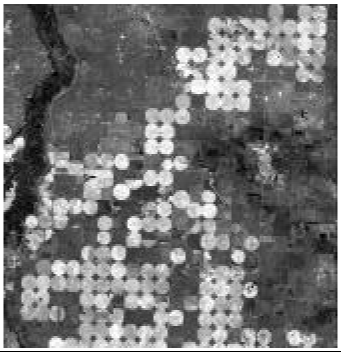
Cover 1998: 7.3%

Are there limitations?

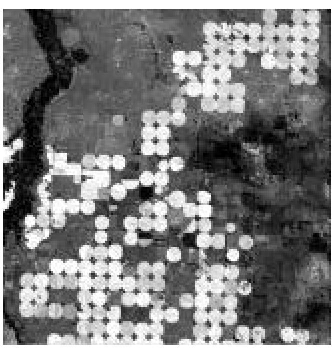
Questions you may ask

- Does it work on other data?
- What is the smallest detectable object size?
- How accurately is crown diameter depicted?
- How separable are objects that are close together?
- How does the wavelet compare to other methods?
- Sensitive to the background?
- What are the omission and commission errors?

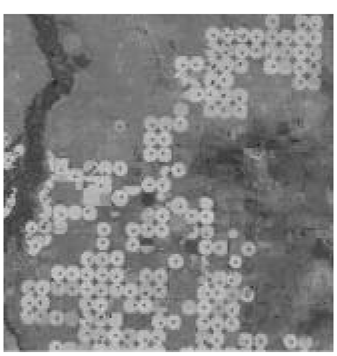
Other Applications:
Pivot Irrigation Crop Circles in Landsat ETM+ Imagery



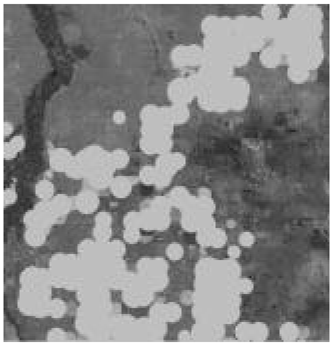
Lets Look with ETM+ Band 7:



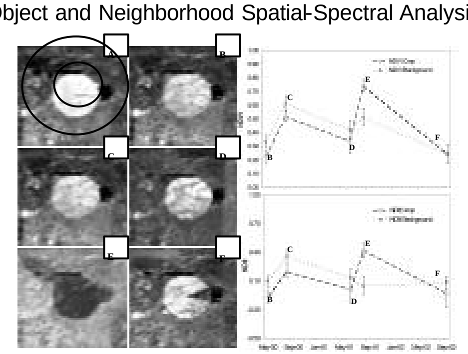
Where are the Objects:



What is the Cover?



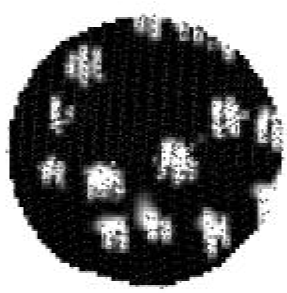
Object and Neighborhood Spatial-Spectral Analysis



Application to LIDAR Height Data:

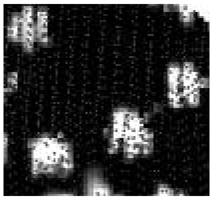
Lidar Point Data (<2m post-spacing) interpolated to 1-m grid of Heights:

Data Info:
Moscow Mountain (ID)
73 Stem-mapped plots
15 selected:
- Simple test case
- Isolated trees N=30

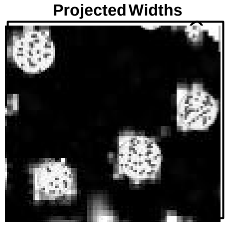


Applying the Method

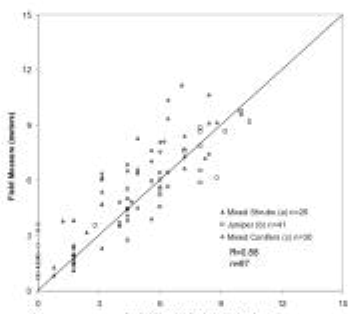
Lidar Height Data



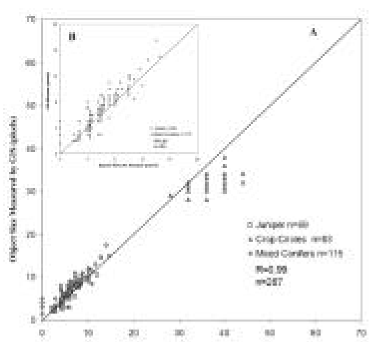
Projected Widths



Wavelets compared to Field measured crown diameter



Wavelets compared to Image measured crown diameter



Detectable size

- The smallest detectable juniper was 2-3 times larger than the image pixel size for both wavelet analysis and digitizing in a GIS
- Bitterbrush or sagebrush did not contribute to commission errors

Wavelet analysis for Object Detection

Points to Note:

Method is:

- Fast
- Repeatable
- Objective
- Insensitive to the background

- Comparison of Wavelet and Field Measured Diameters (R=0.88)
- Comparison of Wavelet and Image Measured Diameters (R=0.99)
- Can detect objects that are 2-3 times larger than the image pixel-size
- Can accurately predict cover in woodlands with cover < 55%
- Can quantitatively analyze both historic aerial photography (i.e. 1939) and modern Remote Sensing datasets

Summary:

Wavelets can be used to detect features that are of the same size and shape (i.e. scale) as the wavelet basis being used.

You can create you own wavelet basis function as long as it is discrete, has a mean of zero, and is broadly similar in shape to the features you are trying to detect.

Several different wavelet techniques exist, but can be broadly divided into using the properties of the:

• Decomposition Scales (i.e. denoising or classification)

• Wavelet Intensities (i.e. feature extraction)

• Wavelet Variances (to determine the scale of your features)