The Mathematics behind JPEG

Jason L. Metcalfe
Johns Hopkins University

February 10, 2003

What is JPEG?

- JPEG = Joint Photographic Experts Group
- JPEG is the industry standard for image compression.
- A digital image is just an array of numbers. Storing such an array takes up a lot of memory…

Types of Image Compression

- Lossless
- The image can always be restored in its full detail.
- Example: GIF (Lempel-Ziv).

- Lossy
- Here we are willing to give up a "little" image quality in order to gain "a lot" of compression.
- Example: JPEG

Lossless Compression

- The idea is to "encode" the data in a different way… abbreviate it.
- Example: If you have a string of a thousand 0's, you might instead store (0,1000).
- Problem: Most strings of data cannot be compressed without loss.

Lossless Compression Still

- In fact, for every "image" that can be compressed by a given algorithm, there will be one that will get larger under that same algorithm.
- Kolmogorov information…
- GIF and other lossless image compression schemes work well on "cartoon" images.

Lossy Compression

- Most lossy compression schemes contain a lossless encoding step. In this talk, we will rather focus on how to represent the "same" image with fewer pieces of data.
- How do we decide which pieces of data to eliminate?

Simplifying Assumptions

- All images are grayscale.
- All images are long and skinny…

- There are ways, of course, to handle color photographs and macroscopic photographs. My assumptions make it easier to work with concrete examples.

Decisions…

- How do we decide which pieces of data to drop? By drop, we mean "set to 0". Using lossless schemes, long strings of zeros can be stored in a small amount of space.
- We want to do so in a way that is "unnoticeable".

Change of Basis

- We need to write our signal in terms of a different basis.
- We will do so in terms of the Fourier basis. This is a basis of pure frequencies.

Euler's formula

$$e^{i\theta} = \cos\theta + i\sin\theta$$

$$i = \sqrt{-1}$$

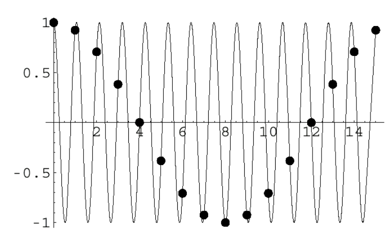
The Fourier Basis

- If N is the dimension of our vector space (e.g., the number of pixels in our image), we want to define N basis vectors: $E_0, E_1, \ldots, E_{N-1}$
- We set the n^{th} component to be

$$E_m(n) = \frac{1}{\sqrt{N}} e^{2\pi i mn/N}$$

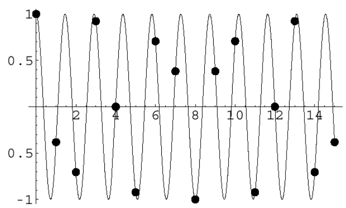
Examples

- If N=16, then the real part of E_{15} is:



Examples

- If N=16, then the real part of E_{11} is



The Fourier Basis

- The Fourier basis vectors consist of pure frequencies.
- When we write our signal (image) in terms of the Fourier basis, the new coefficients represent how much of the corresponding frequency is in the signal.
- This change of basis is often called the Fourier transform.

Fourier Analysis

- A 10-year old's Fourier transform: Open up a grand piano and make a loud bang nearby… what strings will vibrate?
- Fourier analysis (analyzing a signal based on the frequencies that compose it) is used everywhere: radios, televisions, x-rays, telephones,…

Fourier in action…

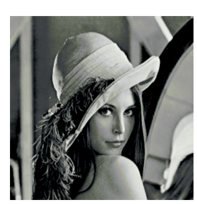
- 1915 – 3 minute coast-to-coast telephone call cost $260 of today's dollars… they could only send one conversation per line. Now, by shifting frequencies, multiple phone conversations are sent on a signal line.

Fourier analysis in JPEG

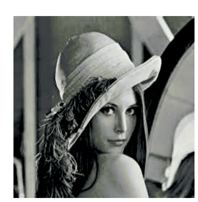
- Most natural images have high spatial correlation.
- Thus, the high frequency components don't contain much visual information. By setting some of these components to zero, we reduce the size of the image without significant loss in quality.

Example: Lena

- BMP – 65 kb



- JPEG, Quality Factor 40 – 6.52 kb



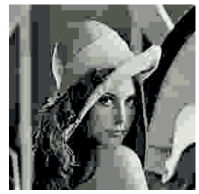
Example: Lena … continued

- JPEG, Quality Factor 20 – 4.35 kb

- JPEG, Quality Factor 10 – 2.97 kb

Example: Lena…

- JPEG, Quality Factor 5 – 2.03 kb



Fast Fourier Transform

- It is also important that the compression scheme can be done quickly.
- Change of basis (vector with N components) – $O(N^2)$
- Fast Fourier Transform – O(N log N)
- FFT is a divide and conquer algorithm… kind of like "Quick Sort".

FFT

- 600 x 400 image –
  - 57,600,000,000 multiplications for change of basis
  - 2,144,721 for FFT
- 3 minute song (7,920,000 samples)-
  - 63,000,000,000,000 for slow FT
  - 90,751,593 for FFT

FFT – last example

- Computing π to a billion digits…
  - $O(N^2)$: ten thousand years
  - O(N log N) : under an hour

JPEG

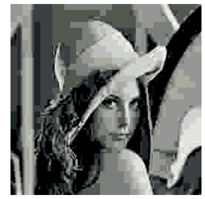
- JPEG uses a related version of the (discrete) Fourier transform called the Discrete Cosine Transform.
- DCT eliminates the need to use complex numbers… and other advantages.
- DCT also has a fast algorithm.

JPEG Disadvantages…

- The Fourier basis vectors are not localized in space. Thus, when you drop a coefficient, it affects the whole image… not just a small part.
- In order to introduce some localization, the image is divided into 8x8 blocks and the Fourier analysis is performed on those 8x8 blocks individually.

JPEG Disadvantages

- There is nothing to guarantee continuity over the boundaries of these blocks. Thus, at high compression ratios, you start to see the blocks.



Wavelets

- The standard basis and the Fourier basis are extremes.
  - The standard basis is well-localized in space, but very poorly localized in frequency.
  - The Fourier basis is well-localized in frequency, but very poorly localized in space.

Wavelets

- We want to find something in between…
- Heisenberg's Uncertainty Principle – you can only do so well. You won't be able to be perfectly localized both in space and frequency.

The Idea: Averages and Details

- Let's take a small concrete example to do be hand:

$$(9,7,3,5)$$

- After one level of averaging and detailing:

$$(8,4,\ |1,\ -1)$$

Iterate the Averaging / Detailing

- Continue the process on the averages…

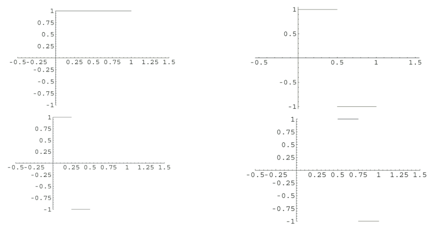
$$(8,4,\ |1,\ -1)$$

- Now becomes

$$(6,|2,\ |1,\ -1)$$

Haar Basis

- What we've really done was write the signal in terms of the basis:



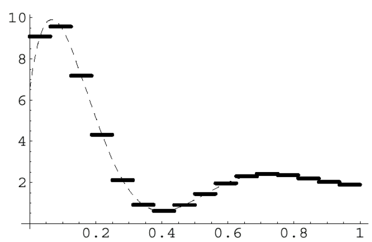
Haar Basis

- In N=4, this is

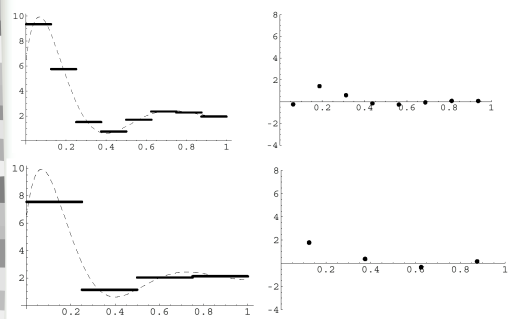
$$\left\{ \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\1\\-1\\-1\end{bmatrix}, \begin{bmatrix}1\\-1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\-1\end{bmatrix} \right\}$$

A More Practical Example

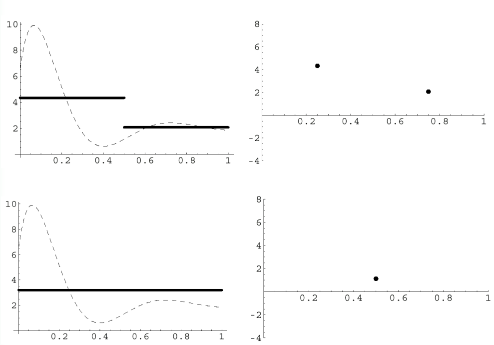
- A sampled signal, 16 samples



Averages: Details



Averages: Details



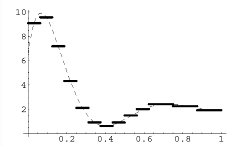
Compression

- Record the final average and the details from every level.
- At this point, we have just changed basis… there is no compression.
- Ideally, however, a number of the detail coefficients will be small… spatial correlation.

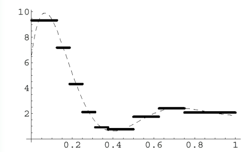
Compression

- To get compression, we “drop” (set to 0) the smallest detail coefficients.
- We can then reconstruct the signal with these missing coefficients… and we get a signal that is approximately the same.

Reconstruction

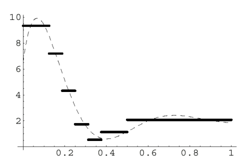


12 coefficients

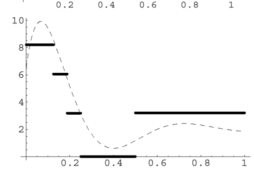


8 coefficients

Reconstruction



6 coefficients



4 coefficients

Wavelet Bases

- The Haar basis tends to be impractical. The “sharp” edges tend to produce noticeable artifacts.
- We want to search for other wavelet bases. That is, other bases that lend themselves to the iterative process that we used in averaging and detailing.

First Stage Wavelet Bases

- Suppose $N=2^n$.
- We are looking for 2 vectors u and v so that

$$\{u(x-2k)\}_{k=0}^{2^{n-1}-1} \cup \{v(x-2k)\}_{k=0}^{2^{n-1}-1}$$

  is an orthonormal basis.
- Here, as we translate, we wrap-around.
- u=father wavelet, v=mother wavelet

Iterative Stage Wavelet Basis

- We then want to iterate as we need with averaging and detailing.
- Here, we want to write the subspace spanned by the translations of u in terms of a similar basis:

$$\{u_2(x-2k)\}_{k=0}^{2^{n-2}-1} \cup \{v_2(x-2k)\}_{k=0}^{2^{n-2}-1}$$

- There is an algorithm that will give this new basis based on the original.

Recursive Algorithm

- How can we find u_2 and v_2 from u_1 and v_1… similarly for later stages:

$$u_2(n) = u_1(n) + u_1\left(n + \frac{N}{2}\right)$$
$$v_2(n) = v_1(n) + v_1\left(n + \frac{N}{2}\right)$$

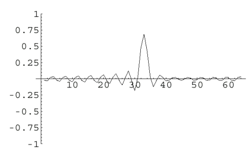
- Where u(n) = n^{th} component of u

Continue the Iteration

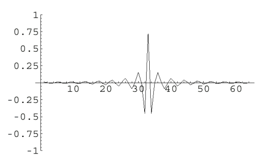
- We then can continue the iteration until the subspace spanned by translations of u is of dimension 1.
- To get compression, we then (as in the case of averaging and detailing) drop the smallest coefficients of the detail basis vectors (the v's).

Real Shannon Wavelets (N=64)

Father Wavelet, u



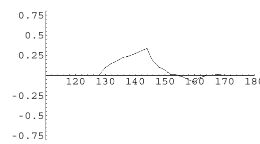
Mother Wavelet v



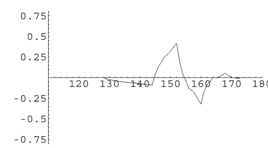
Daubechies D-4 Wavelet: N=256

Father Wavelet, u

(zoomed in)

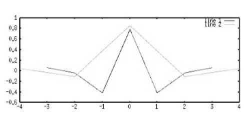


Mother Wavelet, v



JPEG 2000

- Wavelet-based… uses 9:7 wavelet



- A work in progress…

3 MB Original

19 KB JPEG2000

19 KB JPEG

Photo: EE Times

Advantages…

- Built in localization algorithm…
- There is also a fast wavelet transform.
- Built in progressive JPEG…

- From 1-D to 2-D… zig-zag…

Other Applications

- Image querying
- Denoising
- These are just two of many…