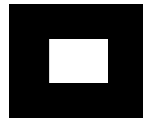
The Mathematics of Image Processing *or* A Picture is Worth a Thousand Equations

Todd Wittman
UCLA Math Circle
April 5, 2009

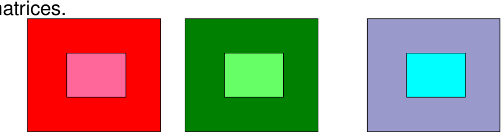
What is an image?

- An image is an integer-valued 2D matrix.
- An 8-bit image takes on values between 0 and 255.

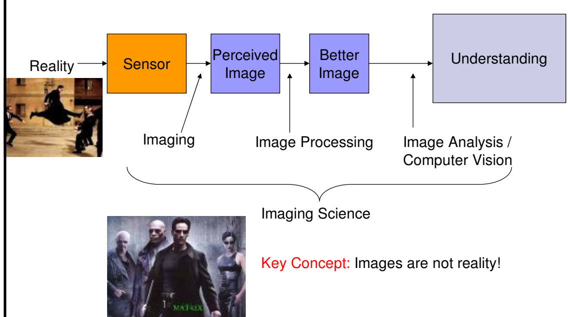


0	0	0	0
0	255	255	0
0	255	255	0
0	0	0	0

- Color images are just processed in three separate 2D matrices.



The Image Pipeline



Key Concept: Images are not reality!

The 9 Basic Tasks of Image Processing

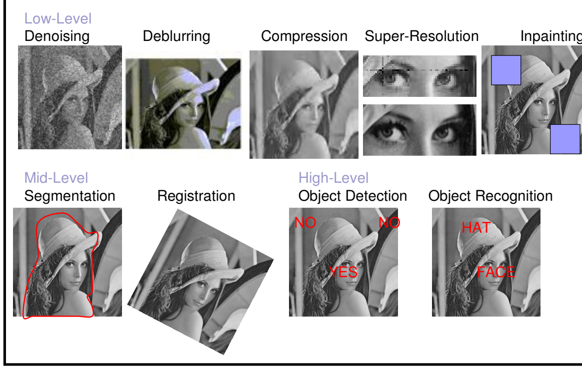


Image Processing at UCLA

- The Mathematics department at UCLA is famous for image processing.
- UCLA developed the *energy-based* approach to image processing, which we'll give you a taste of today.

Andrea Bertozzi | Tony Chan | Stan Osher | Luminita Vese | Jackie Shen

---

Why is image processing hard?

We see this:

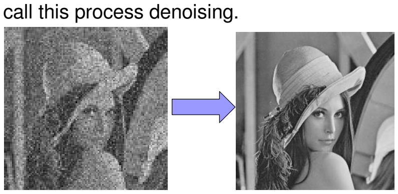
The computer sees this:

207	212	148	2
31	201	253	181
14	255	203	210
26	30	42	45

How do we get the computer to see like us?
We are trying to build a digital soul!

Image Denoising

- We often want to clean up bad images.
- We call this process denoising.

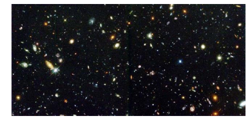


- But the computer doesn't see people, it only sees numbers.
- So mathematically what makes one image better than another?

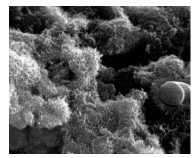
Applications of Denoising

Medical Images

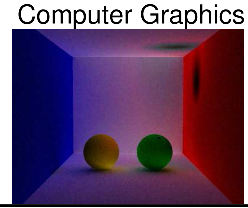
Astronomy



Microscopy



Computer Graphics



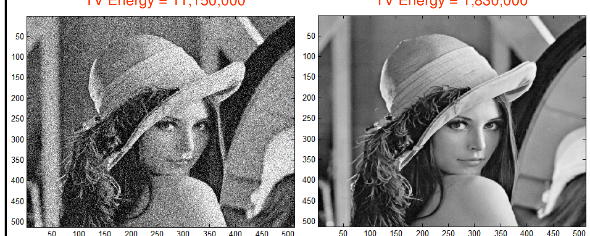
Energy-Based Image Processing

- Physicists say the every system is trying to seek out the lowest energy state (entropy).
- If we can assign some energy function E[u] to an image u:
  - HIGH ENERGY = NOISY
  - LOW ENERGY = CLEAN
- ...then we should drive our image towards the lowest energy.
- There are many different ideas on what the "energy" of an image should be.
- Designing a good energy is the key!

Energy-Based Image Processing

- We are given a noisy, high energy image.
- By minimizing the energy, we remove the noise.

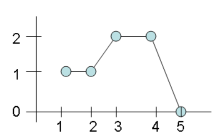
TV Energy = 11,150,000 TV Energy = 1,830,000



Signal Processing

- Images are hard because they are 2-D.
- For starters, let's examine the 1-D case.
- Let's suppose our image is just 1 pixel high.

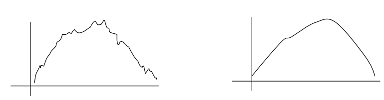
1	1	2	2	0



- We call a 1-D image a *signal*.

The Total Variation (TV) Norm

- What makes one signal noisy and another one clean?



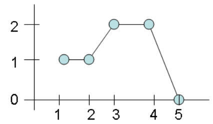
- A popular choice for the energy is the Rudin-Osher-Fatemi Total Variation (TV) energy (1989).
- Define the TV Norm of the signal f(x) to be

$$TV(f) = \sum_{i=2}^{N} |f_i - f_{i-1}|$$

- The signal on the right has smaller TV norm

Example

- Calculate the TV norm.

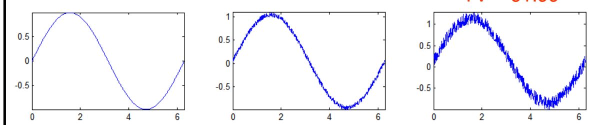


$$\begin{aligned} TV &= |f_2-f_1| + |f_3-f_2| + |f_4-f_3| + |f_5-f_4| \\ &= |1-1| + |2-1| + |2-2| + |0-2| \\ &= 0 + 1 + 0 + 2 \\ &= 3 \end{aligned}$$

TV on Signals

- As we add more noise to a sine wave, the value of the TV norm gets larger.

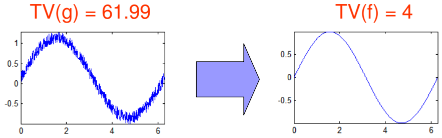
TV = 4 TV = 21.52 TV = 61.99



TV Minimization

- So given a noisy signal *g*, we produce a clean signal *f* by minimizing the TV norm.

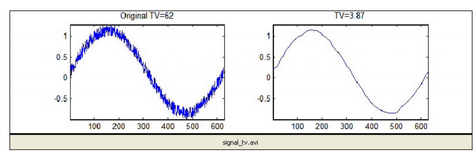
TV(g) = 61.99 TV(f) = 4



- The minimization can be done by the calculus of variations *(which we won't go into)*.
- But what's wrong with just making the TV norm smaller and smaller?

TV Minimization

- If we keep making the TV norm smaller and smaller, we will reach TV=0.



- So the *steady state* of the minimization is a horizontal line.

The Matching Norm

- So we want to minimize the TV norm, but we still want our new signal *f* to resemble the original signal *g*.
- One possibility is to also minimize how well *f* matches *g* in the least squares sense.

$$\min \sum_{i=1}^{N} (f_i - g_i)^2$$

- We call this a *matching* or *fidelity* norm.

Putting It Together

- So how do we minimize the TV norm

$$\min \sum_{i=2}^{N} |f_i - f_{i-1}|$$

- and also minimize the matching norm?

$$\min \sum_{i=1}^{N} (f_i - g_i)^2$$

Putting It Together

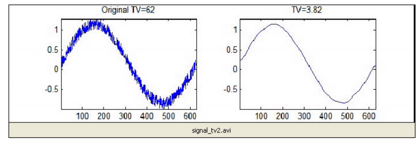
- Answer: Minimize the sum.

$$\min \sum_{i=2}^{N} |f_i - f_{i-1}| + \lambda \sum_{i=1}^{N} (f_i - g_i)^2$$

- This is called the *TV energy*.
- λ is called a *Lagrange multiplier*.
- The parameter controls the balance between the two terms.

TV+Matching Minimization

- When we minimize the TV+matching terms, the final steady state is a clean signal that still resembles the original shape.



The Calculus Version of TV

- The discrete version for a 1-D signal was

$$TV(f) = \sum_{i=2}^{N} |f_i - f_{i-1}|$$

- For a continuous signal, we use a little calculus.

$$TV(f) = \int |f'(x)| dx$$

- Ex Calculate the TV norm of f(x) = sin x on [0,2π].

TV in 2-D

- TV is a sum of the "jumps" with the neighbor.
- This extends to 2-D by looking at the total jump horizontally and vertically.
- For an image u(x,y), define TV to be

$$TV(u) = \sum_{x=2}^{N} \sum_{y=2}^{M} |u(x,y) - u(x-1,y)| + |u(x,y) - u(x,y-1)|$$

Example

- Ex Calculate the TV norm.

0	0	0	0
0	0	1	3
0	0	1	0

TV = 1 + 1 + 1 + 1 +2 + 3 + 3
= 12

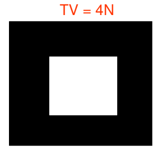
The Isotropic TV Norm

- To make the TV norm rotationally invariant in 2-D, we use the Pythagorean Theorem.

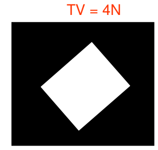
$$TV(u) = \sum_{x=2}^{N} \sum_{y=2}^{M} \sqrt{(u(x,y) - u(x-1,y))^2 + (u(x,y) - u(x,y-1))^2}$$

- Ex We have a NxN white (1) square on a black (0) background. What happens when we rotate it?

TV = 4N

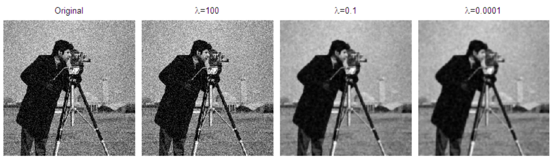


TV = 4N



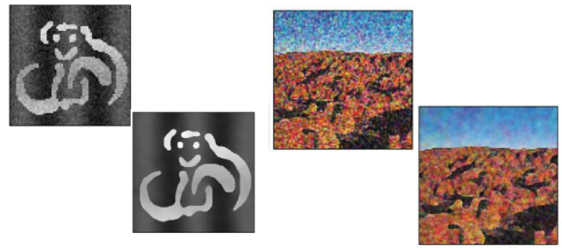
TV Image Denoising

- Adding in the matching term, we can get good denoising results for an appropriate choice of λ.



Color TV Denoising

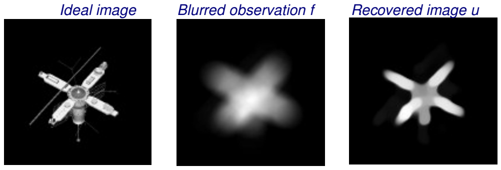
- The TV Energy extends to color images, just minimize each RGB band.



TV Deblurring

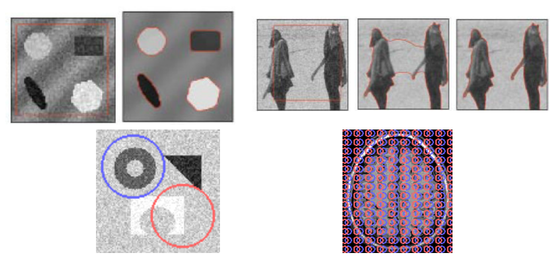
- Blur is another type of problem, different from noise.
- If we have an estimate of the process that caused the blur, TV can remove the blur *(somewhat)*.

Ideal image *Blurred observation f* *Recovered image u*



TV Segmentation

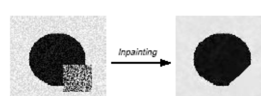
- By looking at the edges that TV emphasizes, we can *segment* the image into pieces.
- This is called Mumford-Shah segmentation.



Slide taken from Andrea Bertozzi, without permission.

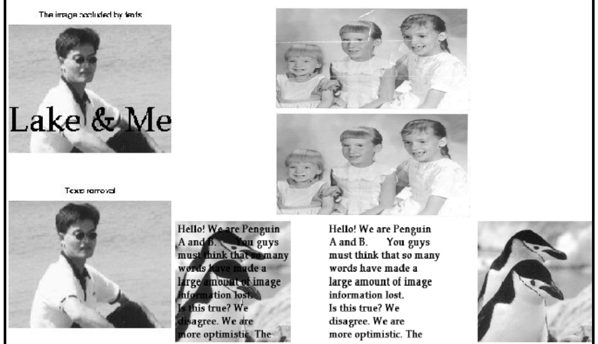
TV Inpainting

- When inpainting a damaged region, we just "turn off" our matching term in the unknown region.
- But this doesn't always complete curves the way we want it to.



- But still, it gives some nice results...

TV Inpainting

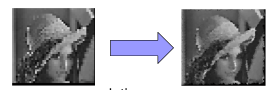


Mumford-Shah-Euler Inpainting by Jackie Shen

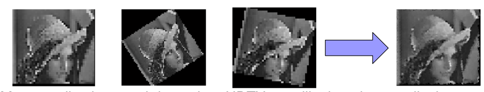
The Super-Resolution Problem

Intuitive Definition: Given a low-resolution image(s), produce an aesthetically pleasing high-resolution image.
Mathematical Definition: ???

•Single-image super-resolution (interpolation)



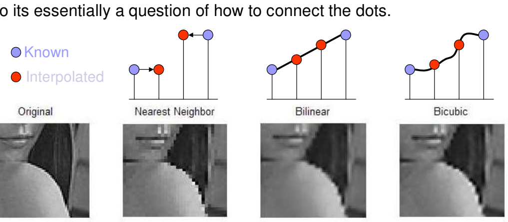
•Multiple-image super-resolution



•Many applications: web browsing, HDTV, satellite imaging, medical diagnosis, surveillance video, recognition

Single Image Interpolation

- Image interpolation (resizing) is essentially filling in pixels in between other pixels.
- So its essentially a question of how to connect the dots.



•The higher degree polynomial we use to interpolate, the better.
•Blurs edges, oversmooths texture, aliasing (staircasing), ringing artifacts.
•May make sense for interpolation of a general data set, but not a good model for *visual* information.

TV Super-Resolution

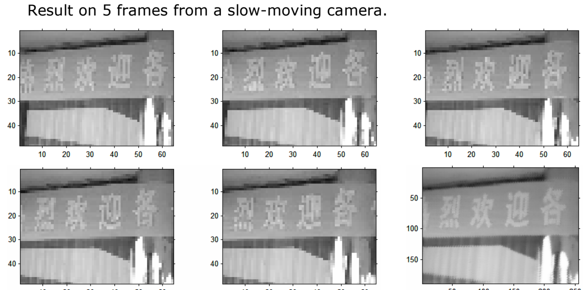
- When we're given multiple images, first line the images up. *(This is actually the hard part.)*

- The produce the image that is smooth (low TV norm) and matches all the images on average.

$$\min_u E\left[u \middle| u_{1\le i\le N}\right] = \int_{\Omega_M} |\nabla u| dx + \frac{\lambda}{2N}\sum_{i=1}^{N} \int_{\Omega_M \cap \varphi_i(\Omega_i)} \left(u - u_i \circ \varphi_i^{-1}\right)^2 dx$$

TV Super-Resolution

Result on 5 frames from a slow-moving camera.



The image gets fuzzier on the ends where there is less information.

Video Super-Resolution

Interlaced traffic video of Karl-Wilhelm-/ Berthold-Straße intersection in Karlsruhe.

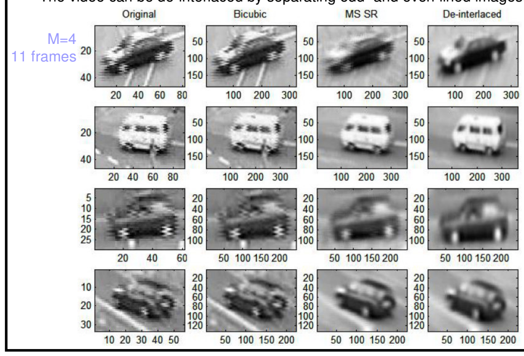
It wouldn't make much sense to super-resolve the whole video.
But we could pull out sections tracking cars, as long as they don't turn the corner.

Video Super-Resolution

It wouldn't make much sense to super-resolve the whole video.
But we could pull out sections tracking cars, as long as they don't turn the corner.

Video Super-Resolution

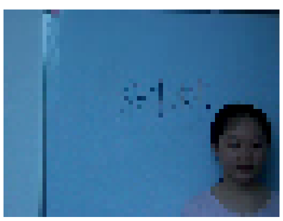
The video can be de-interlaced by separating odd- and even-lined images.



Video Super-Resolution

We can super-resolve a video by aligning the images to each frame.

Low-resolution video



TV super-resolution

That's All Folks!