The Mathematics of Image Processing

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

The Image Pipeline

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.

Why is image processing hard?
We see this:

The computer sees this:

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?

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!

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.
- 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
  \[ TV(f) = \sum_{i=1}^{N} |f_i - f_{i-1}| \]
- The signal on the right has smaller TV norm.
Example

- Calculate the TV norm.

![Example Diagram]

$$TV = |f_2-f_1| + |f_3-f_2| + |f_4-f_3| + |f_5-f_4|$$

$$= |1-1| + |2-1| + |2-2| + |0-2|$$

$$= 0 + 1 + 0 + 2$$

$$= 3$$

TV on Signals

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

![TV on Signals Diagram]

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?

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**.
- \(\lambda\) 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 \([0, 2\pi]\).

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

\[
TV(u) = \sum_{i=2}^{N} \sum_{j=2}^{N} |u(x,y) - u(x-1,y)| + |u(x,y) - u(x,y-1)|
\]

Example

- **Ex** Calculate the TV norm.

\[
\begin{array}{cccc}
0 & 0 & 0 & 0 \\
0 & 0 & 1 & 3 \\
0 & 0 & 1 & 0 \\
\end{array}
\]

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.

\[
TV(u) = \sum_{i=1}^{N} \sum_{j=1}^{N} \sqrt{(u(x,y) - u(x-1,y))^2 + (u(x,y) - u(x,y-1))^2}
\]

- **Ex** We have a \(N\times N\) white (1) square on a black (0) background. What happens when we rotate it?
TV Image Denoising

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

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).

TV Segmentation

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

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...
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 it's 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.

\[
\min \int_\Omega \sum_i \int_\Omega |\nabla u_i|^2 + \lambda \int_\Omega \sum_i \int_\Omega \sum_j (u_j - u_i)^2 \, dx
\]

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.
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.

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

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