CSE 659A: Advances in Computer Vision

Spring 2019: T-R: 2:30-4pm @ Cupples II/230

Instructor: Ayan Chakrabarti (ayan@wustl.edu).

http://www.cse.wustl.edu/~ayan/courses/cse659a/

Apr 11, 2019
Today's the last lecture class.
The next four classes will be project presentations.
Canvas dropbox to submit slides. Due for everyone on Sunday.
- Schedule posted on Piazza and on Canvas.
Prepare for a 12 minute talk + 3 minutes of questions.
Will give you feedback on your proposals by this weekend.
Don't forget about Project 2 and HW 5!
STRUCTURED LIGHT DEPTH SENSORS

- Regular Two-View Stereo

- Have two images of the same scene from different viewpoints.
- Epipolar Geometry reduces search space to along epipolar lines (horizontal lines in rectified stereo).
- If we can find correspondences, then we have depth.
- But finding correspondences is hard. Why?
- Because images are ambiguous, have smooth regions, repeated textures, etc.
- What if we could control what the images looked like?
STRUCTURED LIGHT DEPTH SENSORS

Diagram showing a light source, surface, and two cameras.
STRUCTURED LIGHT DEPTH SENSORS
STRUCTURED LIGHT DEPTH SENSORS

- Basic idea: Cameras and Projectors are duals of each other.
- Camera captures light from a certain scene point at a specific pixel location.
- If instead of a sensor plane, I had a projector image plane: where each "pixel" had different transmittivities based on an "image".
- And I put a lamp behind that plane
- This then projects that image to the world, along the same rays that a camera would have used to capture the image.
• So by replacing one camera with a projector, I can "project" any image I want in the world, record it with the other camera, and look for correspondences between the captured image and projected pattern.

• What should I project for stereo? Images that are "unique" along the epipolar direction.

• Simplest Case:

For a horizontally rectified camera-projector pair, project a vertical line.

In each row, the relative shift in position of the sole bright point gives me disparity.
STRUCTURED LIGHT DEPTH SENSORS

- This is the basis of "laser scanning"

Digital Michaelangelo Project @ Stanford
But remember, this only gives me depth along one line at a time.

To get a full 3D reconstruction, I need to take multiple images while moving the line in the epipolar direction across the image plane.

Can we do this with a single shot? Why not project a full image instead of a line?

Would have to make sure that for each row (same y—co-ordinate), the intensity of the pixel at every x—co-ordinate is unique. (Or perhaps have a repeating pattern based on maximum disparity).

But the problem is that my observed image is not the projected pattern. It is the projected pattern x the underlying image.

Need to be more careful while designing "codes" to project.
STRUCTURED LIGHT DEPTH SENSORS

Horizontal Code repeated Vertically
Some code $c[x]$ for each pixel, which can be used to identify $x$ from $c$ in the observed image $C[x-d(x)]$.

But the problem is that we cannot have arbitrary values in $c[x]$. Because what we observe is $l[x-d] * C[x-d]$

Solution is to have binary codes (or use sinusoidal codes and do correlations across pixels).
But now, we can't distinguish between all the different pixels that have the same value.
So that the code-vector across time for each x is unique

Solution: Take multiple images with different codes. The goal is to maximize depth resolution AND time resolution (small number of images)
STRUCTURED LIGHT DEPTH SENSORS

This type of codes are called phase-shifted codes.

But other general forms of $c(x,t)$ possible.
  e.g., different frequencies

So that the code-vector across time for each $x$ is unique

Solution: Take multiple images with different codes. The goal is to maximize depth resolution AND time resolution
  (small number of images)
STRUCTURE LIGHT DEPTH SENSORS
Because you are taking multiple images of the same scene, you'll observe
\[ I[x - d]C[x - d, 1], I[x - d]C[x - d, 2], I[x - d]C[x - d, 3], \ldots \]

Since the image intensity \( I[x - d] \) is the same, you can also have non-binary codes and normalize by values for the same pixel across the different measurements.

But you have to be careful, because of noise and lack of precision. So the larger the gap between the different values of \( C[x, t] \) the better.

Sometimes, you can also use color. (Though a lot of structured light sensors project and capture in infra-red to be non-intrusive).

Salvi et al., "A state of the art in structured light patterns for surface profilometry," Pattern Recognition 2010

Structured Light Depth was the basis of the Kinect v1 depth sensor.
But beyond multiplication by surface intensity, there are other "light-transport" effects.
Which lead to errors in correspondences and reconstruction.

Researchers have looked at designing optimal codes that are robust to such effects.

Also called, "LIDARs". As the name suggests, it's a RADAR using light waves (instead of radio waves).

The basic principle is, we send light, it bounces off the scene, and returns to our camera, and we want to measure the time it took to come back.

Since we know the speed of light, speed x time = distance. Divide by two to get depth.

But first, how do you measure time of flight? Light travels at $3 \times 10^8 \text{m/s}$. There's no sensor with that kind of temporal accuracy.

Many ways to measure this. But let's talk about the most popular one: using Photon Mixing Devices.

Your projector sends out a pulsing / time-varying light pattern: say $I(t) = \cos(\omega t)$.

The light that comes back is $R(t) = \alpha \cos(\omega(t - \delta))$.

On your receiver, you have two other synchronized signal generators that are producing $C(t) = \cos(\omega t)$ and $S(t) = \sin(\omega t)$.

Photon mixing devices compute temporal correlations of the returning light with these on-chip signals:

$$R_c = \int C(t)R(t)dt, \quad R_s = \int S(t)R(t)dt$$

Look at the phase (or $\tan^{-1} \frac{R_s}{R_c}$) to get $\omega \delta$, and thus $\delta$. 
TIME OF FLIGHT SENSORS

- This is the basic idea, but can use different "modulating" and "de-modulating" functions (shape of light pulse, shape of correlating functions).

- This is the basis of LIDAR (and Kinect V2). Unlike structured light, can work in outdoor scenes better with strong ambient lighting.

Source: Mohit Gupta
TIME OF FLIGHT SENSORS

- In complex scenes, you will again have inter-reflections and multiple light path effects. Can use multiple frequencies and signals to "filter" out inter-reflections.
- Can choose signals so that for each ray sent out, only measure the length of the shortest path back.
- Can separately filter out a list of different path lengths that you get back.
- Can use this for measuring much more than straight depth.
TIME OF FLIGHT SENSORS

- Show light traveling through complex materials.

TIME OF FLIGHT SENSORS

- Use this to "look around corners"

TIME OF FLIGHT SENSORS

- Use to look through dense liquids which scatter light: scattered light paths are longer than direct reflections.

TIME OF FLIGHT SENSORS

- Use to get a full 3D reconstruction of an object (including internal holes) instead of a "depth map", without having to take multiple images from different view-points.

TIME OF FLIGHT SENSORS

- Use to get a full 3D reconstruction of an object (including internal holes) instead of a "depth map", without having to take multiple images from different view-points.

TIME OF FLIGHT SENSORS

- Use to get a full 3D reconstruction of an object (including internal holes) instead of a "depth map", without having to take multiple images from different view-points.

• That brings us to the end of the course! Project presentations in the next two weeks.
• But since this is the first offering of the course: I would appreciate some feedback. Feel free to e-mail me, otherwise enter comments in the course evaluation.
• We covered a lot of topics in this course...
CONCLUSION

Me

( and you after this course )

arXiv
CONCLUSION

What this means for the course

- The course went broad. We tried to cover topics in a diverse set of areas in computer vision.
- But we went into far less depth (compared to CSE 559A).
- It's also infeasible to have 559A-style homeworks when we are covering so many different topics.
- But would you prefer it be done differently? More emphasis on a subset of topics? If so, which ones.
- (Splitting into multiple courses is not an option. We've already split into 559A and 659A).
What this means for the you

If you're going to do research (or practice) in computer vision:

- The rate at which new ideas are coming in and being published is unprecedented.
- Things we talked about in Jan/Feb are probably not state-of-the-art anymore.
- So, you will have to keep reading papers.
- If you're going to be a grad school, a good option is to participate in reading groups and seminars.
- Have your friends tell you about interesting new papers. Tell your friends about interesting new papers.
- Don't restrict yourself to a specific problem type or application domain. Good ideas are transferable.
CONCLUSION

If you want to be a researcher in Vision (or ML)

- When you decide to work on a project, remember the impact of a paper you publish, to your standing, reputation, career, etc. is a highly non-linear function of paper quality!

Source: Bill Freeman, "How to Write a Good CVPR Submission"
What Makes a Good Paper or Project

- Works in practice:
  - Experimental results show improvements in either accuracy or speed (or both!)
  - The results are on "real enough" benchmarks that everyone is convinced they will work "in the wild".
- The Idea is Novel and *Simple*
  - Don't confuse novelty with complexity.
  - Sometimes, you will have to do a bunch of engineering to get the idea to work in practice.
  - But the simpler the idea of your contribution is, the more impact it will have.
  - Easier to convince people that the results are real. You didn't overfit to a benchmark.
  - More likely the idea can be transferred to other applications and domains.
- The best papers maximize performance while minimizing idea complexity.
CONCLUSION

THAT'S ALL FOLKS!

See you for presentations next week!