# RGB-D Sensing

COSC 470: Special Topic
Computer Vision | 3D Reconstruction
Steven Mills

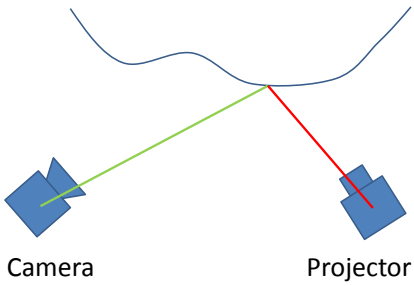
## RGB-D Sensing

- New sensors provide colour (RGB) + depth (D)
  - Microsoft's Kinect most famous
  - Other sensors are available
- Different techniques possible to capture depth
  - Scanning sensors (LiDAR)
  - Time of flight
  - Structured light (Kinect)

## Structured Light Basics

- Stereo vision relies on matching points
- If we replace one camera with a projector
  - We can project a ray into the scene
  - This lights up a single point in the world
  - We can easily see that point in the camera image
  - Triangulation is then used to find the 3D point

## Structured Light Basics




## Structured Light

- This gives us a single point
- We could scan that over the scene – slow
- We can project a regular grid of points
  - Assumes that the left-right and top-bottom ordering is preserved
  - Projecting a set of lines makes this ordering clearer – could use a grid for example

## Structured Light in the Kinect

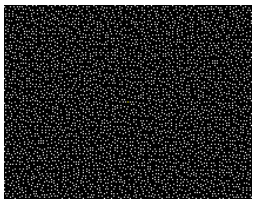
- The Kinect has three main components that concern us
  - A visible light camera
  - An infrared projector
  - An infrared camera
- Details of structured light method a bit hazy
  - Proprietary approach developed by PrimeSense
  - Also uses depth from focus

## The Kinect Sensor


Infra-Red Projector | Colour Camera | Infra-Red Camera

## Kinect – Basic Structured Light

- Kinect projects an infra-red speckle pattern
- This creates a pattern of dots that the IR camera can see
- But this would give only a sparse set of 3D points, and not very accurate



http://azttm.wordpress.com/2011/04/03/kinect-pattern-uncovered/

## Depth from Focus

- If you focus on a near object, then distant objects are blurred
- The more distant the object, the more the blur
- You can measure blurriness in images – gradients get lower
- Kinect uses “astigmatic” lenses – different focal lengths in x- and y- directions

## Depth From Focus

- This setup causes the dots to blur into ellipses
- The size and angle of the ellipses depends on the depth
- This allows a dense, accurate depth map to be reconstructed

## Working with the Kinect

- Microsoft provides an SDK for Windows
  - Fairly high level interface
  - Can get access to the depth and colour maps
  - Can find corresponding points in the two maps
  - Skeleton tracking
  - Audio features – detecting direction to noises
  - Can tilt the sensor up and down, put it into ‘near mode’, change resolution, etc.

## Demo and Code (I hope)

## Kinect Fusion

- By combining multiple depth maps you can make a 3D scanner
- Can't run it on my laptop (GPU accelerated)
- So YouTube:

  http://www.youtube.com/watch?v=quGhaggn3cQ
  http://www.youtube.com/watch?v=NsrmniEvO4s

## Combining Point Clouds

- Kinect Fusion merges multiple point clouds
- To do this it uses Iterative Closest Point (ICP)
  - Start with roughly aligned point clouds
  - Find nearest-neighbour matches between them
  - Each match gives us a vector which estimates how to move one cloud to align it with the other
  - Compute a rotation and translation from these vectors – similar to transforms for normalised DLT

## Skeleton Tracking

- The Kinect does 'skeleton tracking'
  - Fits a stick-figure model to the depth map
  - Tracks this over time
  - Useful for gesture recognition
  - Used as a natural user interface
- Also has a microphone array – can triangulate sounds to identify people making noise etc.

## Using Multiple Kinects

- You can use up to 4 Kinects in one PC
- You can have a network of them
- However, their structured light can interfere
  - Can overcome this with a vibrating motor (eg: from a mobile phone) on each Kinect
  - Vibrations cause the projected pattern to blur
  - Each Kinect's camera and projector move together – no blur observed

## RGB-D Advantages

- Gives you direct depth estimates
- Doesn't rely on textured scenes or lighting
- Gives a dense, regular sampling of depth
- Cheap
- No calibration required (done in factory)
- Gives you colour + depth in a single package

## RGB-D Disadvantages

- Active sensor – projects light into the scene
- Limited range – a few metres
- Can interfere with each other
- Issues with bright sunlight (IR light source)
- Limited resolution (640x480, 1080p soon)
- Limited depth accuracy (a few mm at best)