CSCI 473/573
Human-Centered Robotics
Almost all modern robots have color-depth sensing capability!
Stereo Depth Sensing
How Stereo Vision Works

Real life object
Whether it is possible to evaluate depth using one eye?
How Stereo Vision Works

Real life object

Left Camera

Stereo Vision

Right Camera
How Stereo Vision Works

Real life object

Every pixel in an image defines a 3D ray
If you identify a real world point in both images...
How Stereo Vision Works

Every pixel in an image defines a 3D ray
If you identify a real world point in both images...
...You can find the 3D position by triangulation
How Stereo Vision Works

Real life object

$\Delta$centeroid = $(dl + dr)$

$\frac{\Delta$centeroid}{$f$} = \frac{b}{d}$

$f = \text{focal length}$

Common Centroid

Lens Focal Length

Object Distance

Image Distance

Left Camera

Stereo Vision

Right Camera

L lens

R lens

L detector

R detector
How Stereo Vision Works

Camera Calibration

- To know the 3D rays, the cameras must be calibrated!
- Calibration also corrects lens distortion

Distorted image before calibration

Corrected image after calibration
How Stereo Vision Works

Camera Calibration

- To know the 3D rays, the cameras must be calibrated!
- Calibration also corrects lens distortion

Distorted image before calibration
Corrected image after calibration

Wow - that's distorted!
Big difference!
Straight lines are straight again
How Stereo Vision Works

Stereo Matching

- Now we need to match pixels in the images

Left Image

Right Image
How Stereo Vision Works

Stereo Matching

- Now we need to match pixels in the images
- To do that, take a region around a given pixel

Left Image | Right Image
When camera is leveled, it is sufficient to only search the horizontal line.
Point Grey
Innovation in Imaging

How Stereo Vision Works

Stereo Correlation

• Find matches for every pixel and you can create a depth image
How Stereo Vision Works

Stereo Correlation

- Find matches for every pixel and you can create a depth image

Depth Image

Hot pixels are close to the camera
Cool pixels are farther away
Two representations of depth, but not completely equivalent
Stereo Depth Sensing

• Two main steps:
  • Correspondence (more challenging)
  • Reconstruction
Stereo Depth Sensing

• Failure of Correspondence search

Another issue: SLOW!
Stereo Depth Sensing

Stereo Vision based indoor/outdoor Navigation for Flying Robots

Autonomous quadrotor flight using waypoints given by an operator. On-board stereo image processing, sensor data fusion, mapping and path planning. Visualization of on-board data.

For this video, map resolution was increased off-board from 0.1m to 0.02m.
Big Dog by Boston Dynamics

http://www.youtube.com/watch?v=cNZPRsrwumQ
Structured-light 3D Sensing

• Color-depth camera

Less than $100

~ $150

Bought by Apple Inc. for $350 million in Nov, 2013
Kinect
Other Kinect Hacks

http://www.youtube.com/watch?v=31ebpITAvss
How Kinect works?

- Microsoft licensed this technology from PrimeSense
- The depth computation is all done by the hardware built into Kinect
- Details are not completely publicly available (not available in 2014)
# Kinect 1 vs Kinect 2

<table>
<thead>
<tr>
<th>Feature</th>
<th>Kinect for Windows 1</th>
<th>Kinect for Windows 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color Camera</td>
<td>640 x 480 @30 fps</td>
<td>1920 x 1080 @30 fps</td>
</tr>
<tr>
<td>Depth Camera</td>
<td>320 x 240</td>
<td>512 x 424</td>
</tr>
<tr>
<td>Max Depth Distance</td>
<td>~4.5 M</td>
<td>~4.5 M</td>
</tr>
<tr>
<td>Min Depth Distance</td>
<td>40 cm in near mode</td>
<td>50 cm</td>
</tr>
<tr>
<td>Horizontal Field of View</td>
<td>57 degrees</td>
<td>70 degrees</td>
</tr>
<tr>
<td>Vertical Field of View</td>
<td>43 degrees</td>
<td>60 degrees</td>
</tr>
<tr>
<td>Tilt Motor</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Skeleton Joints Defined</td>
<td>20 joints</td>
<td>26 joints</td>
</tr>
<tr>
<td>Full Skeletons Tracked</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>USB Standard</td>
<td>2.0</td>
<td>3.0</td>
</tr>
</tbody>
</table>

*Image via [blogs.msdn.com](http://blogs.msdn.com)*
Kinect v1

Kinect v2 prerelease
How Kinect Works

One viewpoint sometimes cannot work well.
How Kinect Works

Stereo sensing is OK in some (or most) cases.
How Kinect Works

But... NOT all.
How Kinect Works

IR projects an invisible pattern to the scene
How Kinect Works

The pattern can differentiate linear view changes
How Kinect Works

The pattern can be also used to locate the same point, thus solving the correspondence problem.
How Kinect Works

The pattern can be also used to locate the same point, thus solving the correspondence problem.
How Kinect Works

So... What is the key technology in Kinect?
How Kinect Works

So... What is the key technology in Kinect?

Projected Pattern!
How Kinect Works

• Must be different between each other *(to address the correspondence challenge)*
• Must be able to encode depth *(to better and faster estimate depth)*
How Kinect Works

Projected Pattern!

• Must be different between each other
  -- **Structured Light**
• Must be able to encode depth
Correspondence using Structured Light

Structured light general principle:

Project a known pattern onto the scene and measure the similarity of the sensed and projected patterns to find the correspondence

Correspondence using Structured Light

Recall of stereo vision:
• Hard to solve the correspondence problem
• Slow

- Correspondence problem
- 3D reconstruction of matched pairs by triangulation
Correspondence using Structured Light

Color-encoded structured light:

- One of the cameras is replaced by a light emitter
- Correspondence problem is solved by searching the pattern in the camera image (pattern decoding)
Correspondence using Structured Light

Early color-encoded structured light:

- Use color lines as the projected pattern
- Need a more complex correspondence algorithm

Works despite complex appearances

Works in real-time and on dynamic scenes

Correspondence using Structured Light

Structured light general principle:
• Kinect uses infrared light with a speckle pattern.
• Color is not used for computing depth.

Unfortunately, we do not know the complete design detail.

Shpunt et al, PrimeSense patent application, US 2008/0106746
Correspondence using Structured Light

Structured light general principle:

- No pattern groups look like any another
- The distinctive points make matching easy.
Correspondence using Structured Light

Region-growing Random Dot Matching

1. Detect dots (“speckles”) and label them unknown
2. Randomly select a region anchor (a dot with unknown depth)
   a. Windowed search along a scan line
      Check that best match score is greater than threshold; if not, mark as “invalid” and go to 2
   b. Region growing
      1. Neighboring pixels are added to a queue
      2. For each pixel in queue, initialize by anchor’s shift; then search small local neighborhood; if matched, add neighbors to queue
      3. Stop when no pixels are left in the queue
3. Repeat until all dots are matched or are marked “invalid”

Depth Sensing Using Structured Light

Projected Pattern!

• Must be different between each other
  -- Structured Light
• Must be able to encode depth
  -- Depth from focus
Depth from focus

Principle:
Stuff that is more blurry is further away

• The Kinect dramatically improves the depth estimation accuracy based stereo vision
• The Kinect uses a special lens with different focal length in x and y-directions
• A projected circle then becomes an ellipse whose orientation depends on depth (distance)
Depth from focus

Principle:
Projected circle becomes an ellipse whose orientation depends on depth

Freedman et al, PrimeSense patent application, US 2010/0290698
Depth from focus

Kinect projected pattern change
Depth from focus

Kinect projected pattern change
Depth from focus

Kinect projected pattern change
Projected circle becomes an ellipse whose orientation depends on depth (but also more blurry when depth increases)

Example: Book vs. No Book
Depth from focus
Depth Sensing using Kinect

Projected Pattern!

• Must be different between each other
  -- Structured Light
• Must be able to encode depth
  -- Depth from focus
  -- Depth from stereo
Depth from Stereo

- Same stereo depth sensing algorithms directly apply
Depth from Stereo

• But it solves the correspondence issues and image distortion (to some extent).
Depth Sensing using Structured Light

The Kinect combines
-- **structured light**
with two classic computer vision techniques:
-- **depth from focus**
-- **depth from stereo**

The KEY is the pattern design!
Projected IR vs. Natural Light Stereo

• What are the advantages of IR?
  – Works in low light conditions
  – Does not rely on having textured objects
  – Not confused by repeated scene textures
  – Can tailor algorithm to produced pattern

• What are advantages of natural light?
  – Works outside, anywhere with sufficient light
  – Uses less energy
  – Resolution limited only by sensors, not projector

• Difficulties with both
  – Very dark surfaces may not reflect enough light
  – Specular reflection in mirrors or metal causes trouble
Simultaneous Localization And Mapping

http://www.youtube.com/watch?v=9Y4RQVpp-BY
Human-Robot Interaction
Human-Robot Interaction

Use kinect to detect and track humans. Get human position to avoid humans.