Sensing the World with Kinect
A Parade of 3D Applications

Ceyhun Burak Akgül, PhD
Vistek ISRA Vision & Boğaziçi University EE Dept.
www.cba-research.com
Kinect is a motion sensing device by Microsoft originally developed for the Xbox 360 video game console and Windows PCs.

It enables users to control and interact with the Xbox 360
• without the need to touch a game controller,
• through a natural user interface using gestures and spoken commands.
In a bit more detail...
Google search with “Kinect”: 60 Million results!
Facts and Figures

• Launched in North America on November 4, 2010

• *Fastest selling consumer electronics device* - according to Microsoft
  – 8 million units sold in its first 60 days.
  – 24 million units of the Kinect sensor shipped by February 2013
  – An advertising budget of US$500 million for the launch

• Depth sensing technology due to Prime Sense

• "Soft" technologies due to Microsoft Research
  (Kinect for Windows SDK)
  – People/gesture tracking and recognition
  – Facial recognition
  – Voice recognition
What is Kinect capable of sensing?

- Depth map of objects in a scene
- Color image (RGB video)
- Sound
What is Kinect capable of sensing?

- Depth map of objects in a scene
- Color image (RGB video)
- Sound

Color Image  Depth Map  Segmentation Map
Depth Sensing Principle – 1/2

Stereo Reconstruction (Passive)

Ideal Case

Actual Case
Depth Sensing Principle – 2/2

Reconstruction w/ Structured Light (Active)

Projector – Camera Geometry

Structured Light Pattern
## Kinect Specs

### Kinect 1.0

<table>
<thead>
<tr>
<th>Field of View (FOV)</th>
<th>57.5° horizontal by 43.5° vertical</th>
<th>70° horizontal by 60° vertical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolvable Depth</td>
<td>0.8 m -&gt; 4.0 m</td>
<td>0.8 m -&gt; 4.0 m</td>
</tr>
<tr>
<td>Color Stream</td>
<td>640 x 480 x 24 bpp 4:3 RGB @ 30fps</td>
<td>1920 x 1080 x 16 bpp 16:9 YUY2 @ 30 fps</td>
</tr>
<tr>
<td>Depth Stream</td>
<td>320 x 240 16 bpp, 13-bit depth</td>
<td>512 x 424 x 16 bpp, 13-bit depth</td>
</tr>
</tbody>
</table>

Spatial x/y resolution = 3mm (@ 2m distance from sensor)
Depth z resolution = 1cm (@ 2m distance from sensor)
What can intelligent software do with Kinect?

Capture and understand motion

Recognize people by their face

Recognize what people say

Xbox Features
What else can be done with Kinect?

APPLICATIONS
Mixed / Augmented Reality
Mouse-free Design
Semantic Scene Understanding

Depth Image Streams

Human Motion Capture
Gesture-Based Control
Avatar Animation
Kinematic Evaluation
Action Recognition

3D Structure
3D Scanning / Reconstruction
Robot Guidance / SLAM
Object Recognition
Control by Gesture
Control by Gesture
Avatar Animation
Avatar Animation
“Avatar” Animation
Object Recognition
Robot Grasping / Guidance
3D Scanning / Reconstruction
Mouse-free Design
What’s next?
Capturing 3D with a mobile?
It’s already there!

PrimeSense demonstrates Capri 3D sensor on Nexus 10 (hands-on)
What’s next next?