UTS:CAS

CENTRE FOR AUTONOMOUS SYSTEMS

41014 Sensors and Control for Mechatronic Systems

Dr. **Liang Zhao**

Centre for Autonomous Systems
University of Technology,
Sydney

UNIVERSITY OF TECHNOLOGY SYDNEY

0.2 Lecture-2

UNIVERSITY OF TECHNOLOGY SYDNEY

- **Lecture:**
  - Introduction to different sensors
  - Camera: Geometry
  - Camera: Calibration
  - Image Processing: Convolution
- **Active hands on:**
  - Camera Calibration Toolbox
  - Convolution with different Kernels

41014 Sensors and Control for Mechatronic Systems

Lecture-3: RGB-D Cameras

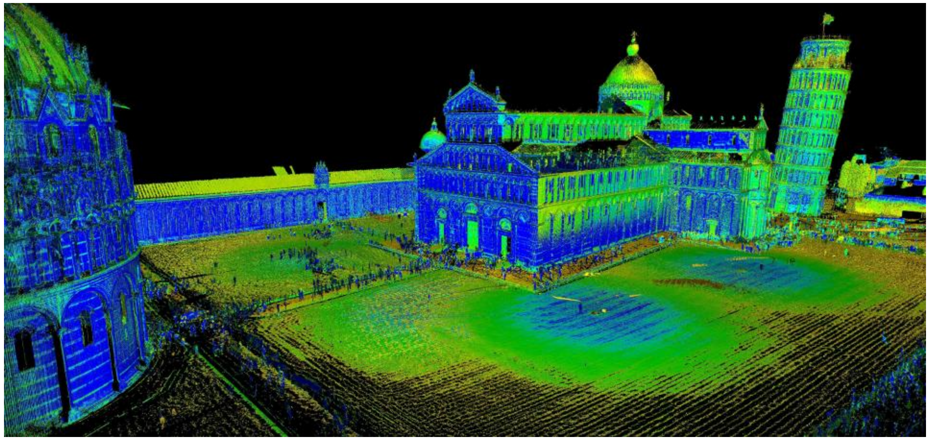
Dr. **Liang Zhao**

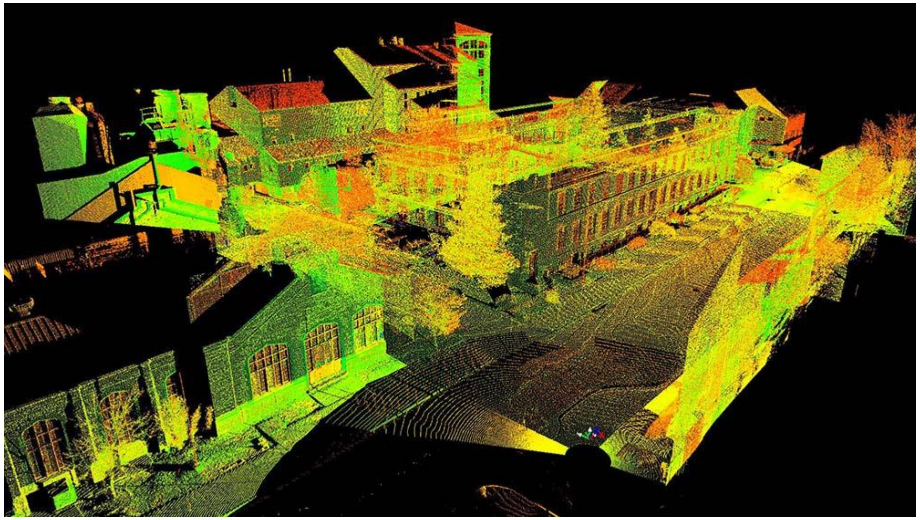
Centre for Autonomous Systems
University of Technology, Sydney

1. Lecture-3

- **Sensors:**
  - **Cameras**
  - **RGB-D Cameras**
  - ToF sensors

- **In details of**
  - Fundamental
  - Data and Processing





1. Lecture-3

UNIVERSITY OF TECHNOLOGY SYDNEY

Lecture:

- Stereo Cameras
- RGB-D Cameras: principle and different models
- Applications of RGB-D Cameras

Active hands on:

- Play with Different RGB-D Cameras
- Display of RGB/Depth images and point clouds
- Image processing on RGB-D images

2.1 Stereo Cameras

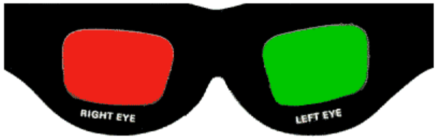
UNIVERSITY OF TECHNOLOGY SYDNEY

Human eyes

Stereo vision

Anaglyphs

- Two images of complementary color are overlaid to generate one image.
- Glasses required (e.g. red/green)
- Each eye gets one image => 3D impression



2.1 Stereo Cameras

Shutter glasses

- Display flickers between left and right image (i.e. each even frame shows left image, each odd frame shows right image)
- Uses a timing signal to synchronize the glass and the display
- Requires new displays of high frame rate (120Hz).

Shutter glasses and 120Hz display

Autostereoscopic displays

- No glasses required!
- Matrix of many transparent lenses put on the display.
- Lenses distort pixels so that left eye gets a left image and right eye gets a right image (if you are standing in a proper spot) => 3D impression

Sharp 3D display

2.2 Stereo: Disparity

- Disparity: the pixel difference in two images

2.2 Stereo: Disparity

- Disparity: the pixel difference in two images

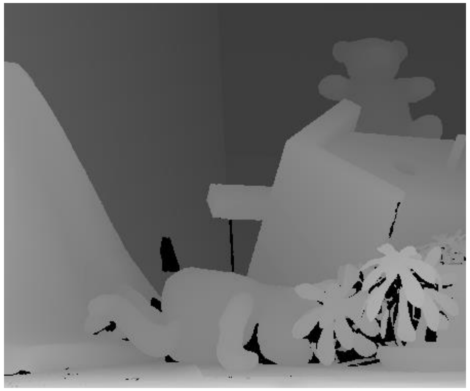
Foreground disparity - large

Background disparity - small

2.2 Stereo: Disparity

❖ Disparity

- The disparity of each pixel is encoded by a grey value.
- High grey values represent high disparities (and low gray values small disparities).
- The resulting image is called disparity map.
- The disparity map contains sufficient information about the depth map

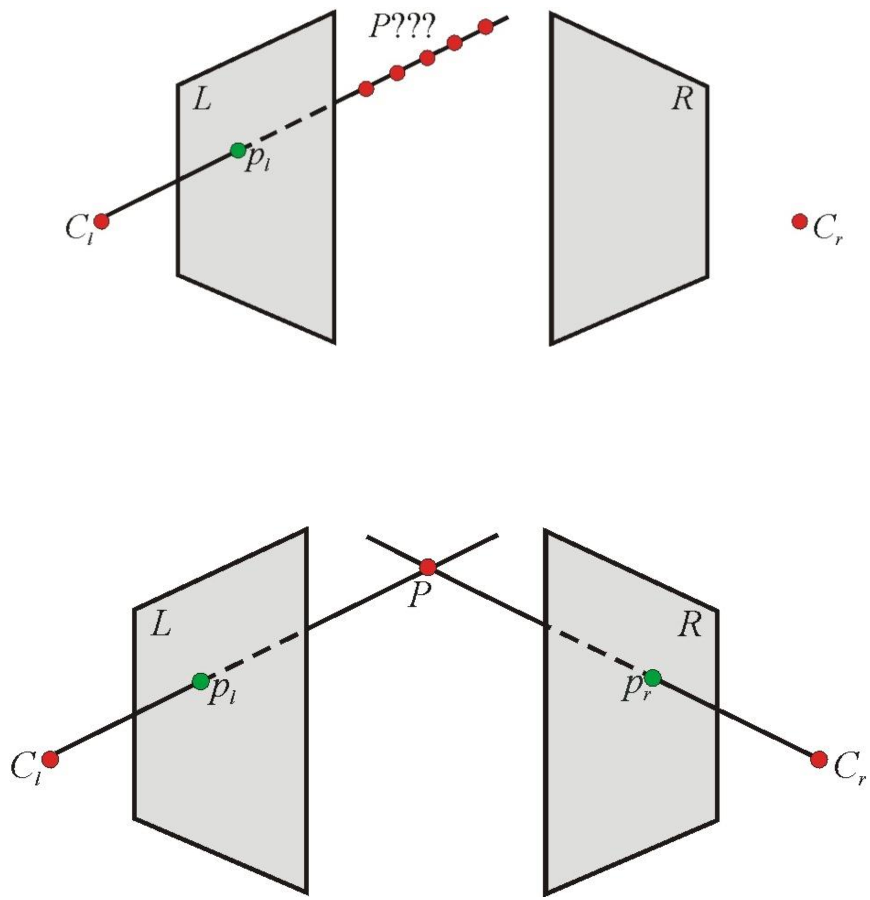


2.3 Stereo: 3D Reconstruction

UNIVERSITY OF TECHNOLOGY SYDNEY

3D Reconstruction

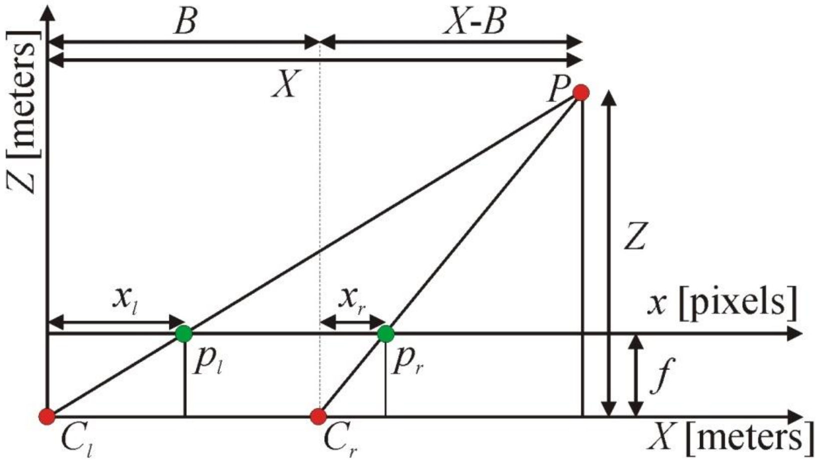
- P can lie on anywhere along the C_l-p_l line.
- Let us assume we also know the 2D projection p_r of P onto the right image plane R. Can we now determine the pint P?
- The challenge is to find the matching p_l and p_r. It is called the stereo matching (correspondence) problem.



2.3 Stereo: 3D Reconstruction

❖ Triangulation

- X, Z : 3D space
- B – baseline of the cameras
- x_l, x_r : Image coordinates
- f : focal length



- Using similar triangles

$$\frac{X}{Z} = \frac{x_l}{f}$$

$$\frac{X - B}{Z} = \frac{x_r}{f}$$

2.3 Stereo: 3D Reconstruction

UNIVERSITY OF TECHNOLOGY SYDNEY

❖ Triangulation

$$\frac{X}{Z} = \frac{x_l}{f} \qquad \frac{X - B}{Z} = \frac{x_r}{f}$$

$$X = \frac{Z.x_l}{f} \qquad X = \frac{Z.x_r}{f} + B$$

$$\frac{Z.x_l}{f} = \frac{Z.x_r}{f} + B$$

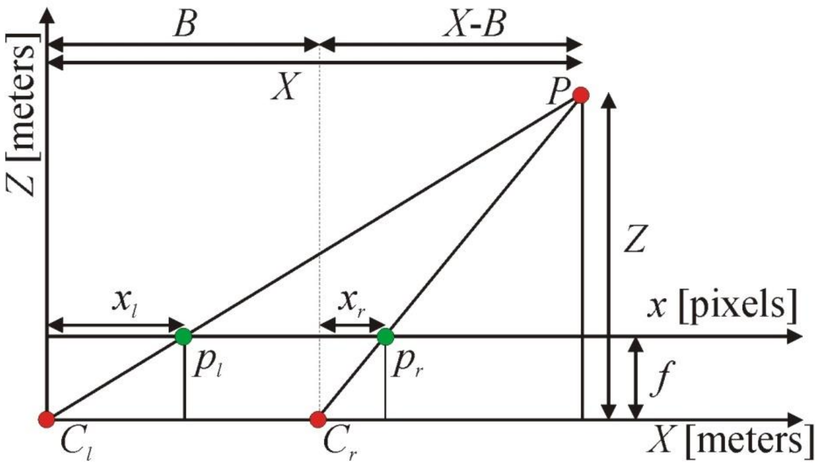
$$Z.x_l = Z.x_r + B.f$$

$$Z.(x_l - x_r) = B.f$$

$$Z = \frac{B.f}{x_l - x_r} = \frac{B.f}{d}$$

Where *d* is the disparity

– It can be seen that the disparity is inversely proportional to the depth.

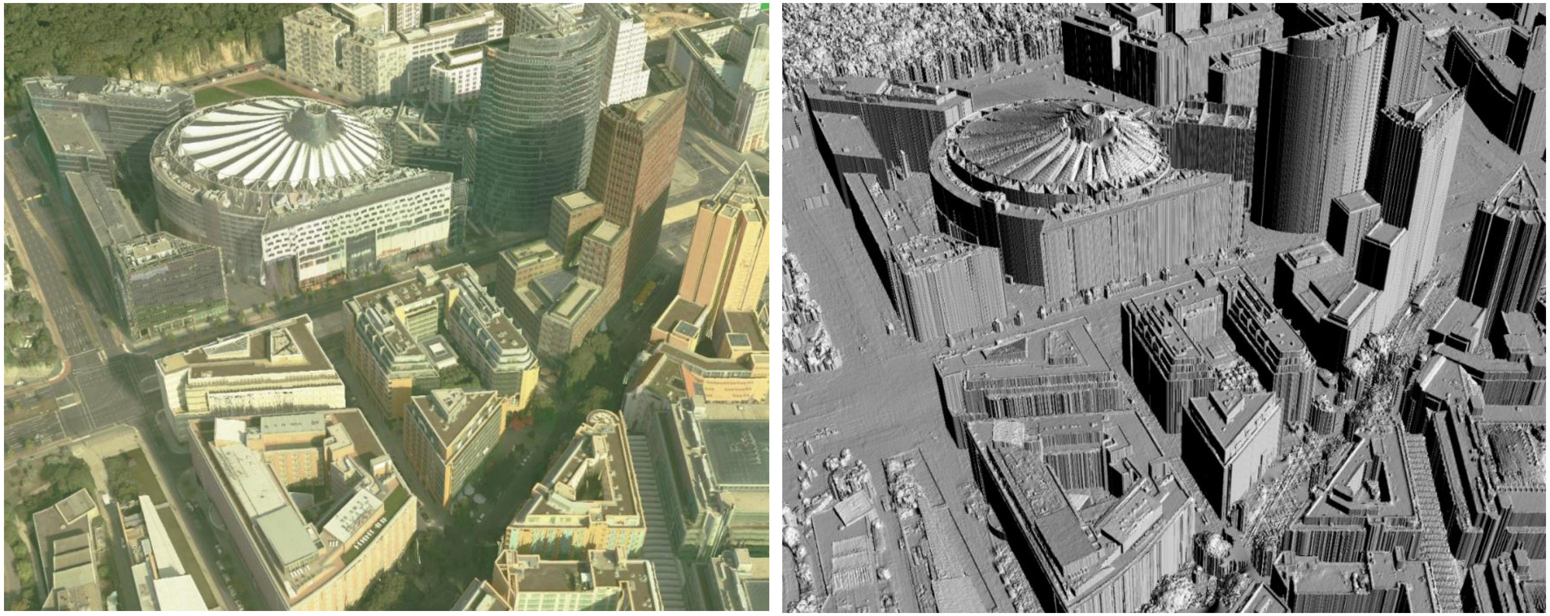


2.3 Stereo: Geometry

Activity 1

- Image resulotion: 800*600
- Principle point: (400,300)
- Focal length: 800
- Baseline: 0.2m
- Image point left (300,300)
- Image point right (200,300)
- 3D Point in left camera frame?

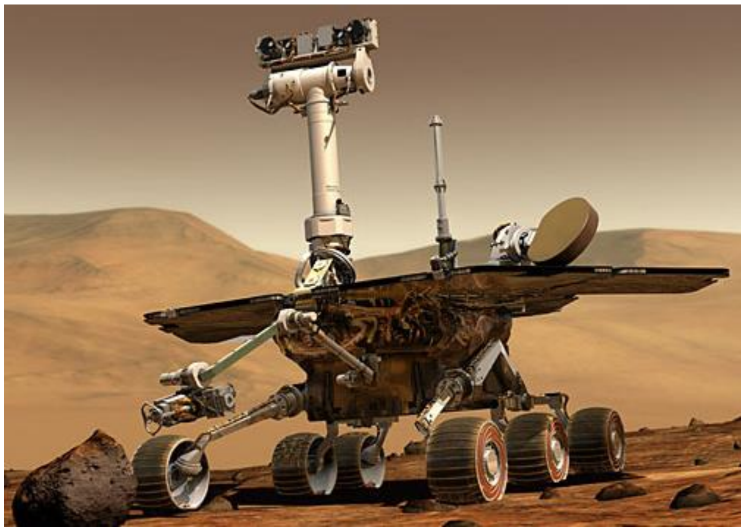
2.5 Stereo: Applications

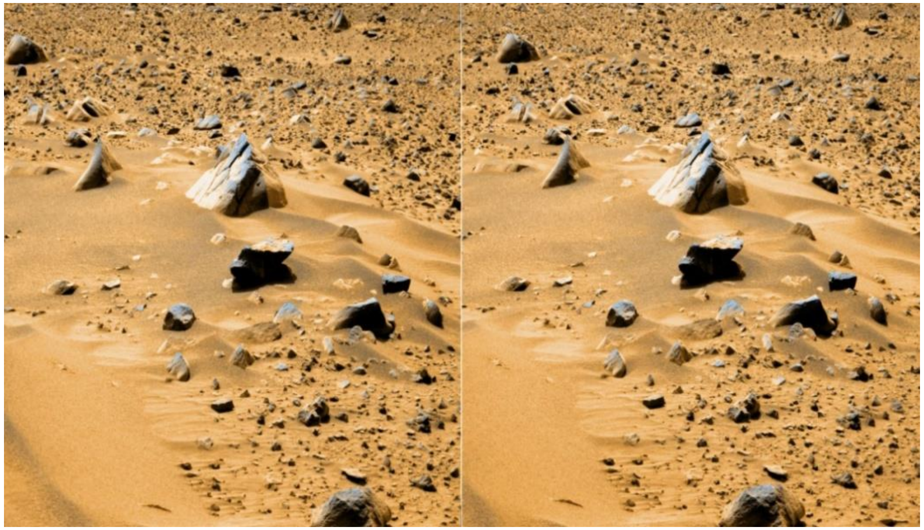


- Stereo cameras are mounted on an airplane to obtain a terrain map.

http://www.robotic.de/Heiko.Hirschmueller/

2.5 Stereo: Applications





- Reconstruct the surface of Mars using stereo vision

2.4 Stereo: Challenges

❖ Stereo Correspondence - challenges

- **Color inconsistencies:**
- When solving the stereo matching problem, we typically assume that corresponding pixels have the same intensity/color (= Photo consistency assumption)
- It is not necessarily be always true:
  - Image noise
  - Different illumination conditions in left and right images
  - Different sensor characteristics of the two cameras.
  - Specula reflections (mirroring)
  - Sampling artifacts

2.4 Stereo: Challenges

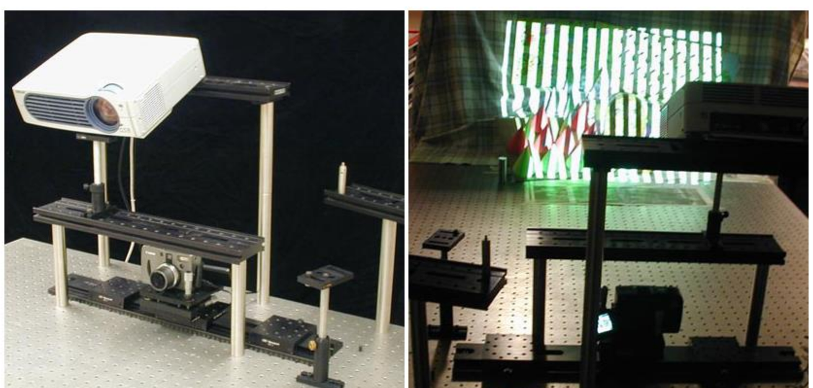
Stereo Correspondence - challenges

- Textureless regions

Left image

Right image

- The corresponding is infeasible. So disparities can not be calculated.



Disparity image

2.4 Stereo: Challenges

Stereo Correspondence - challenges

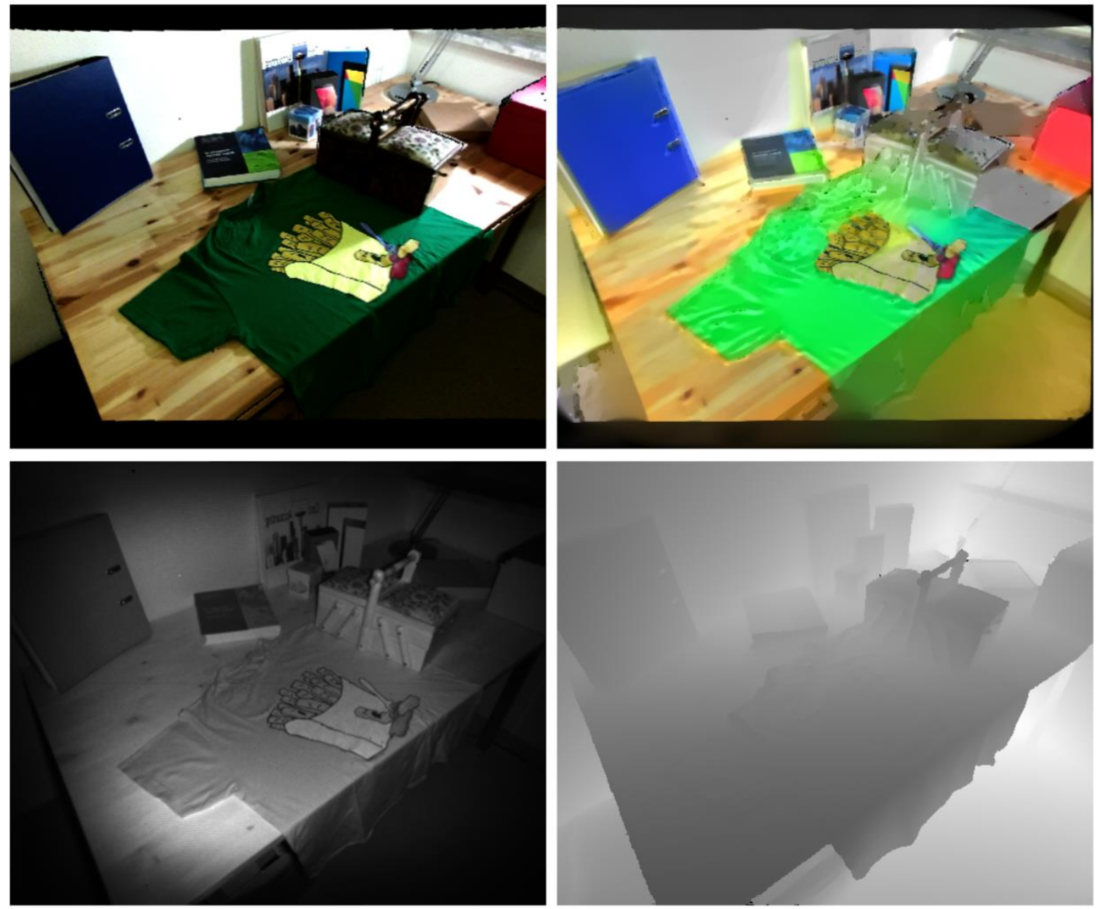
- **Occlusion Problem**
  - There are pixels that are only visible in exactly one view.
  - We call this pixels occluded (or half-occluded)
  - It is difficult to estimate depth for these pixels.
  - Occlusion problem makes stereo more challenging than a lot of other computer vision problems

Occluded pixels

3.1 RGB-D Cameras

❖ RGB-D Cameras:

- Normal RGB Image
- Depth Image

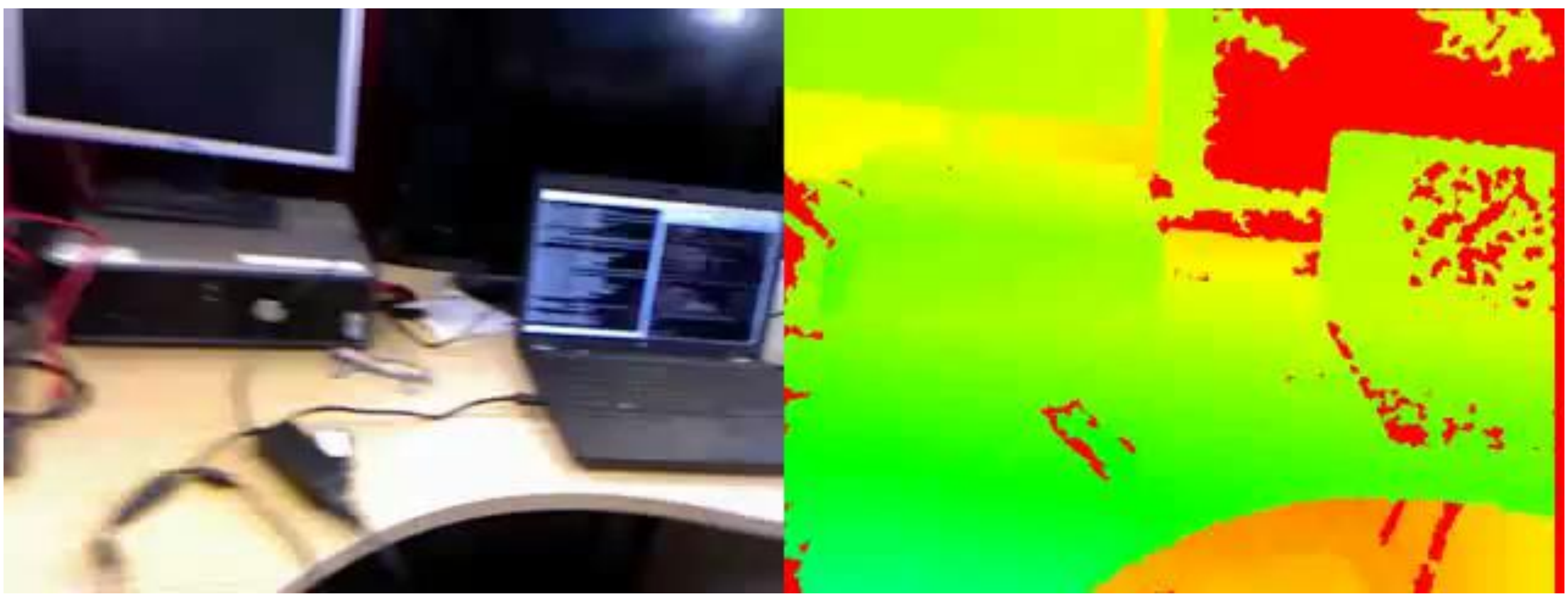


3.1 RGB-D Cameras

UNIVERSITY OF TECHNOLOGY SYDNEY

❖ RGB-D Cameras:

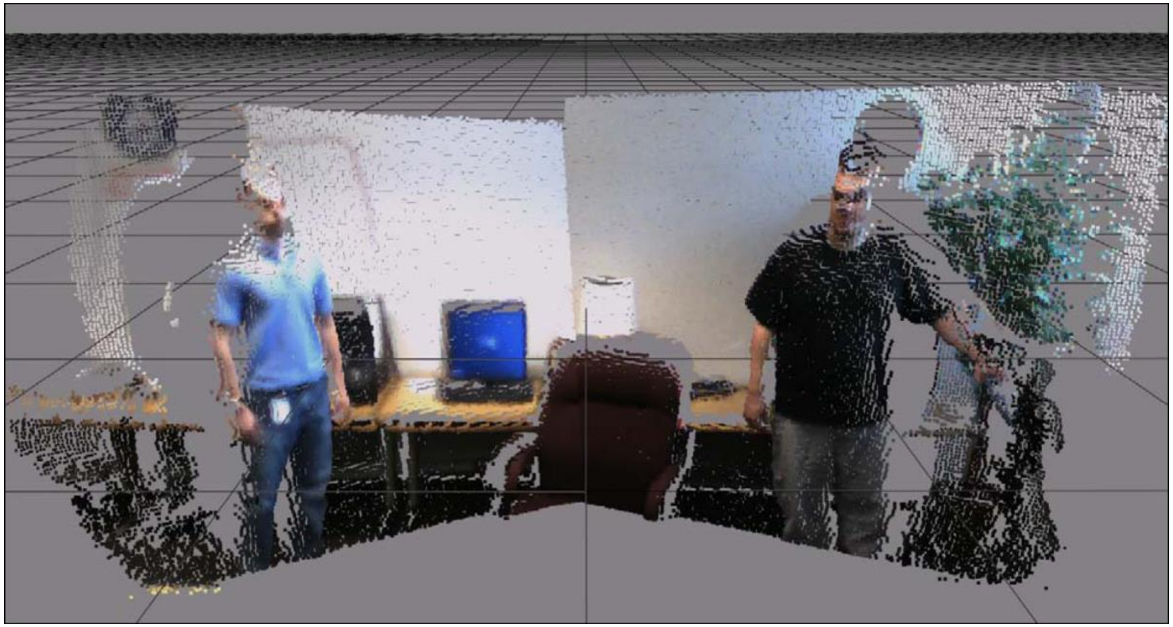
- Normal RGB/Depth Image



3.1 RGB-D Cameras

- RGB-D Cameras

- SwissRanger
  - Time-of-Flight (ToF)
  - Color Camera
  - $10K



3.2 RGB-D: Principle

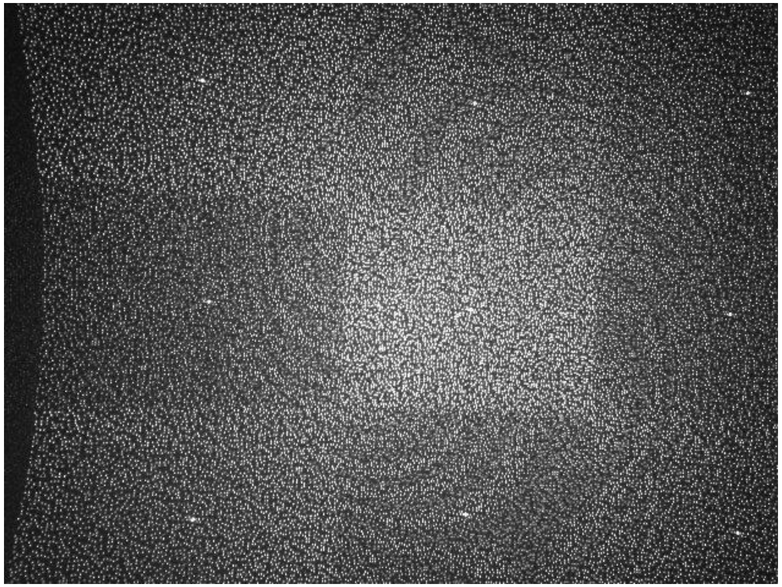
❖ RGB-D Cameras: Kinect

- $200-300
- An Infrared projector
- A color camera
- An Infrared sensor

3.2 RGB-D: Principle

❖ RGB-D Cameras: Kinect

- **Structured Light:**
- Projects an infrared speckle pattern
- The projected pattern is then captured by an infrared camera in the sensor
- Compared part-by-part to reference patterns stored in the device.
- These patterns were captured previously at known depths.
- The sensor then estimates the per-pixel depth based on which reference patterns the projected pattern matches best.

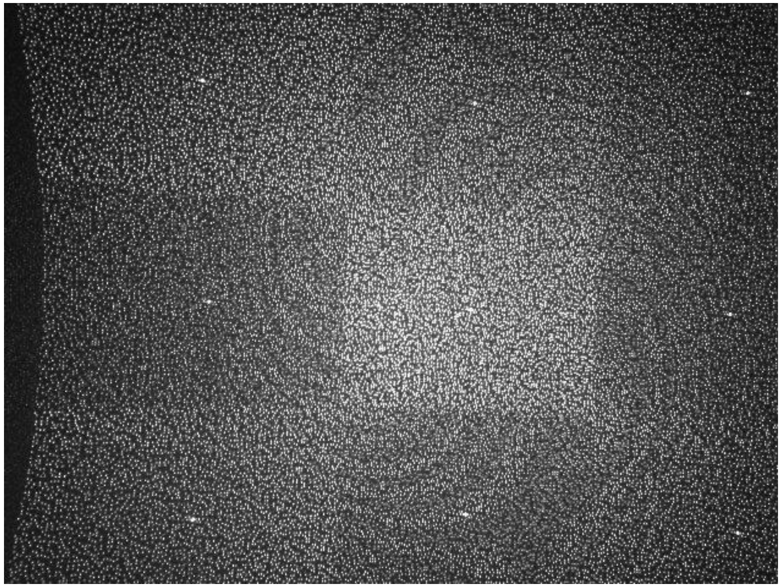


3.2 RGB-D: Principle

UNIVERSITY OF TECHNOLOGY SYDNEY

❖ RGB-D Cameras: Kinect

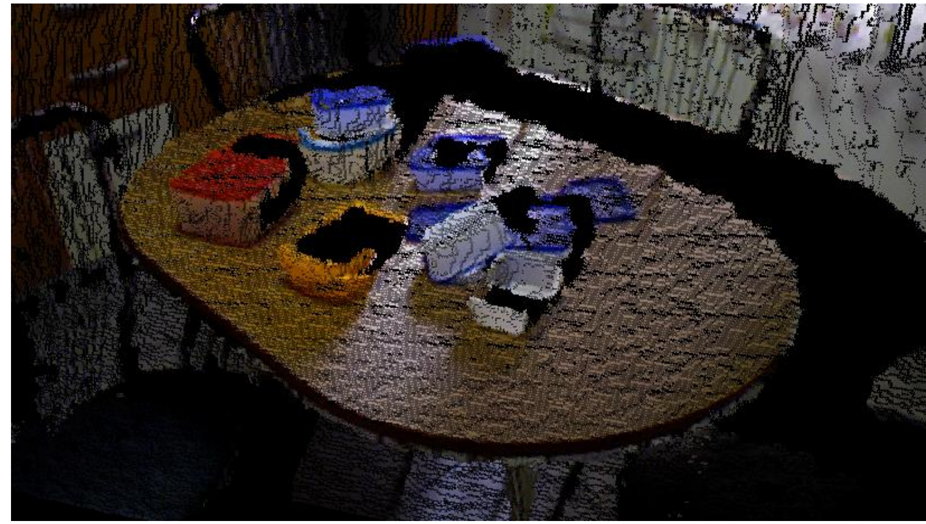
- **Associated to RGB image:**
- The depth data provided by the infrared sensor is then correlated to a calibrated RGB camera.
- This yields an RGB image with a depth associated with each pixel.



3.2 RGB-D: Principle

❖ RGB-D Cameras: Kinect

- **Point Cloud :**
- a collection of points in three dimensional space, where each point can have additional features associated with it.
- With an RGB-D sensor, the color can be one such feature.
- Approximated surface normals are also often stored with each point in a point cloud.



3.3 RGB-D: Different Products

❖ Microsoft Kinect

Distance of Use:
Between 0.5m and 4.5m

Field of view	43° vertical by 57° horizontal field of view
Frame rate (depth and color stream)	30 frames per second (FPS)
Default resolution, depth stream	VGA (640 x 480)
Default resolution, color stream	VGA (640 x 480)
Audio format	16-kHz, 16-bit mono pulse code modulation (PCM)
Audio input characteristics	A four-microphone array with 24-bit analog-to-digital converter (ADC) and Kinect-resident signal processing, such as acoustic echo cancellation and noise suppression

3.3 RGB-D: Different Products

❖ Microsoft Kinect2

Feature	Kinect for Windows 1	Kinect for Windows 2
Color Camera	640 x 480 @30 fps	1920 x 1080 @30 fps
Depth Camera	320 x 240	512 x 424
Max Depth Distance	~4.5 M	8 M
Min Depth Distance	40 cm in near mode	50 cm
Depth Horizontal Field of View	57 degrees	70 degrees
Depth Vertical Field of View	43 degrees	60 degrees
Tilt Motor	yes	no
Skeleton Joints Defined	20 joints	25 joints
Full Skeletons Tracked	2	6
USB Standard	2.0	3.0
Supported OS	Win 7, Win 8	Win 8
Price	$249	$199

3.3 RGB-D: Different Products

UNIVERSITY OF TECHNOLOGY SYDNEY

❖ Asus Xtion PRO

Distance of Use:
Between 0.8m and 3.5m

Field of View	58° H, 45° V, 70° D
Depth Image Size	VGA (640x480) : 30 fps QVGA (320x240): 60 fps
Resolution	SXGA (1280*1024)
Interface	USB2.0

3.3 RGB-D: Different Products

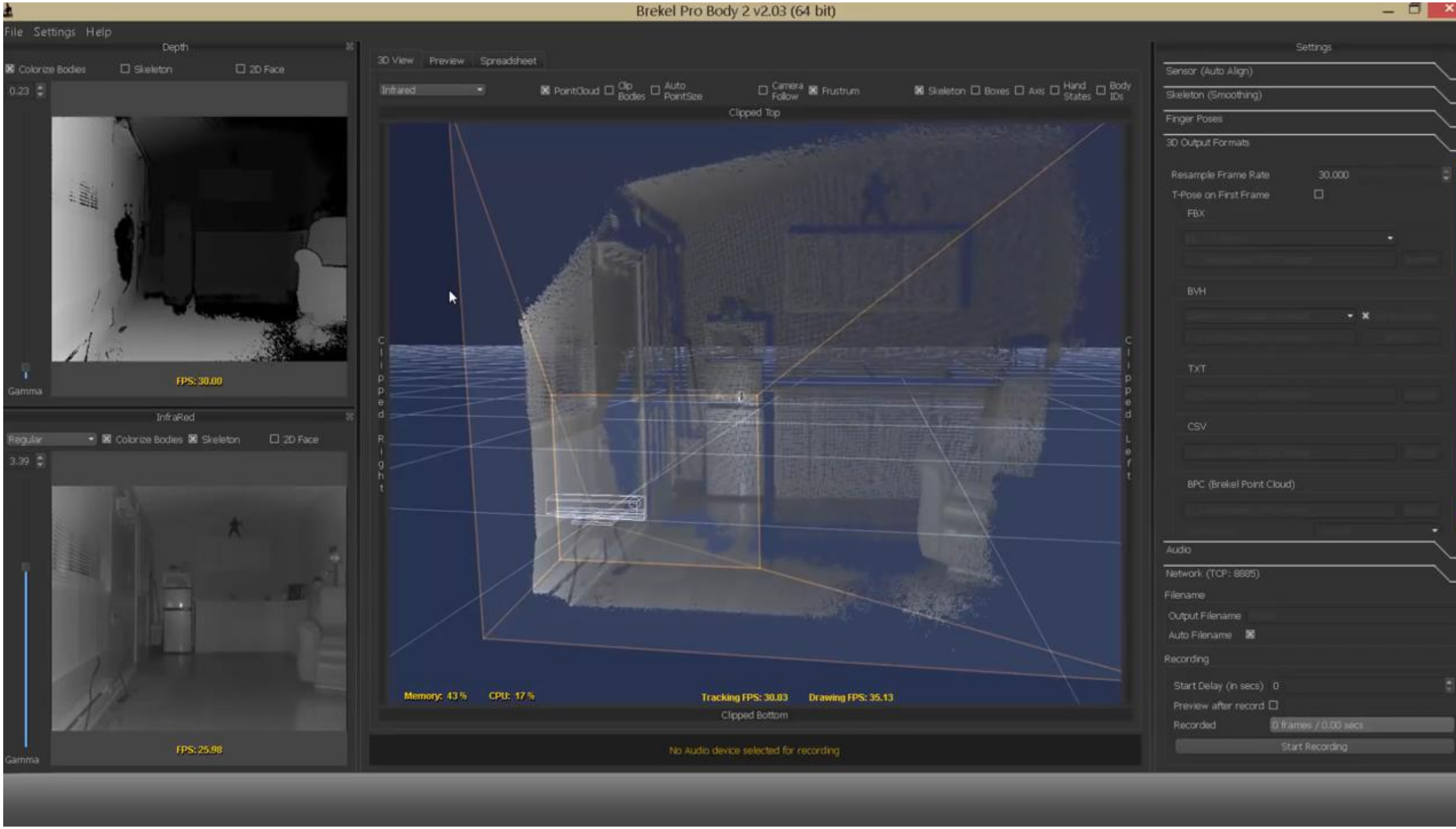
Creative BlasterX Senz3D

- RGB video resolution
  HD 720p (1280x720)
- IR depth resolution
  QVGA (320x240)
- Frame rate: Up to 30 fps
- FOV (Field-of-View): 74°
- Range: 0.5ft ~ 3.25ft
- Dual-array microphones
- Size: 4.27” x 2.03” x 2.11”
- Weight: 271g
- Multi-attach base
- Cable length: 1.8 meters
- USB 2.0 Hi-Speed

3.4 RGB-D: Applications

UNIVERSITY OF TECHNOLOGY SYDNEY

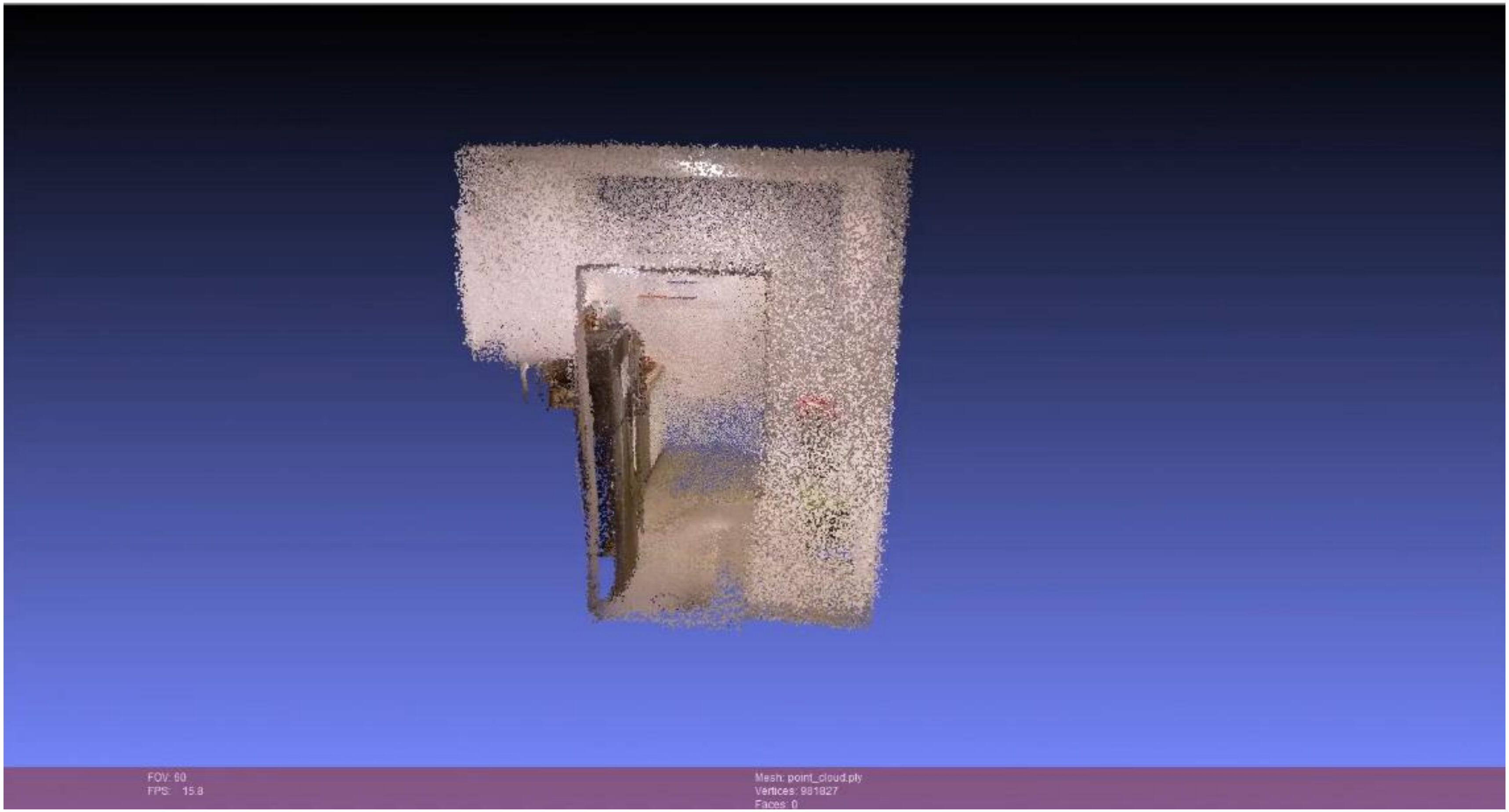
❖ Real-time Skeletal Tracking



3.4 RGB-D: Applications

UNIVERSITY OF TECHNOLOGY SYDNEY

❖ 3D Reconstruction



3.4 RGB-D: Applications

UNIVERSITY OF TECHNOLOGY SYDNEY

- **KinectFusion**

SIGGRAPH Talks 2011

KinectFusion:

Real-Time Dynamic 3D Surface Reconstruction and Interaction

Shahram Izadi 1, Richard Newcombe 2, David Kim 1,3, Otmar Hilliges 1, David Molyneaux 1,4, Pushmeet Kohli 1, Jamie Shotton 1, Steve Hodges 1, Dustin Freeman 5, Andrew Davison 2, Andrew Fitzgibbon 1

1 Microsoft Research Cambridge 2 Imperial College London
3 Newcastle University 4 Lancaster University
5 University of Toronto

DynamicFusion

*Dynamic*Fusion:

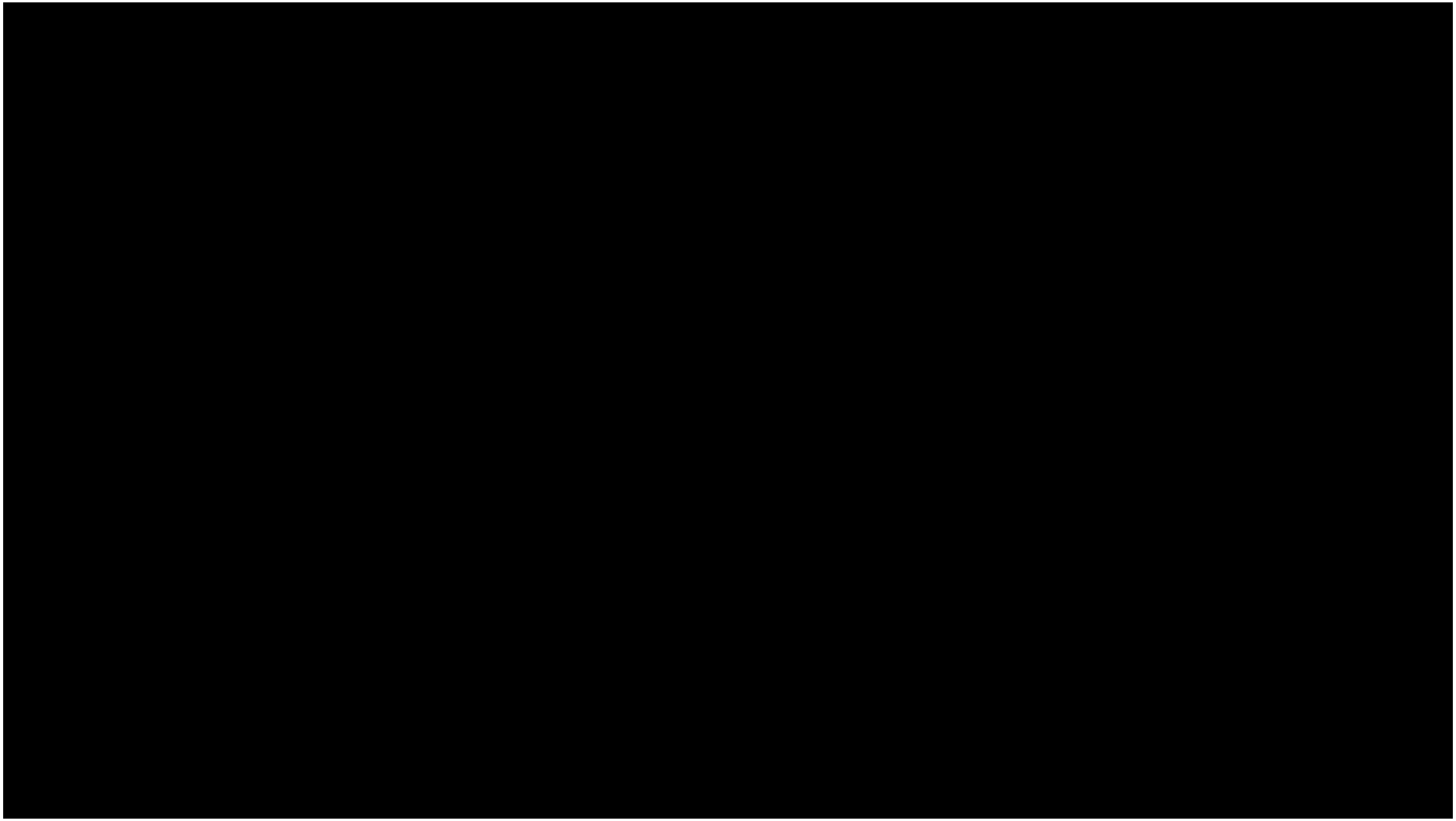
Reconstruction & Tracking of Non-rigid Scenes in *Real-Time*

Richard Newcombe, Dieter Fox, Steve Seitz

Computer Science and Engineering,
University of Washington

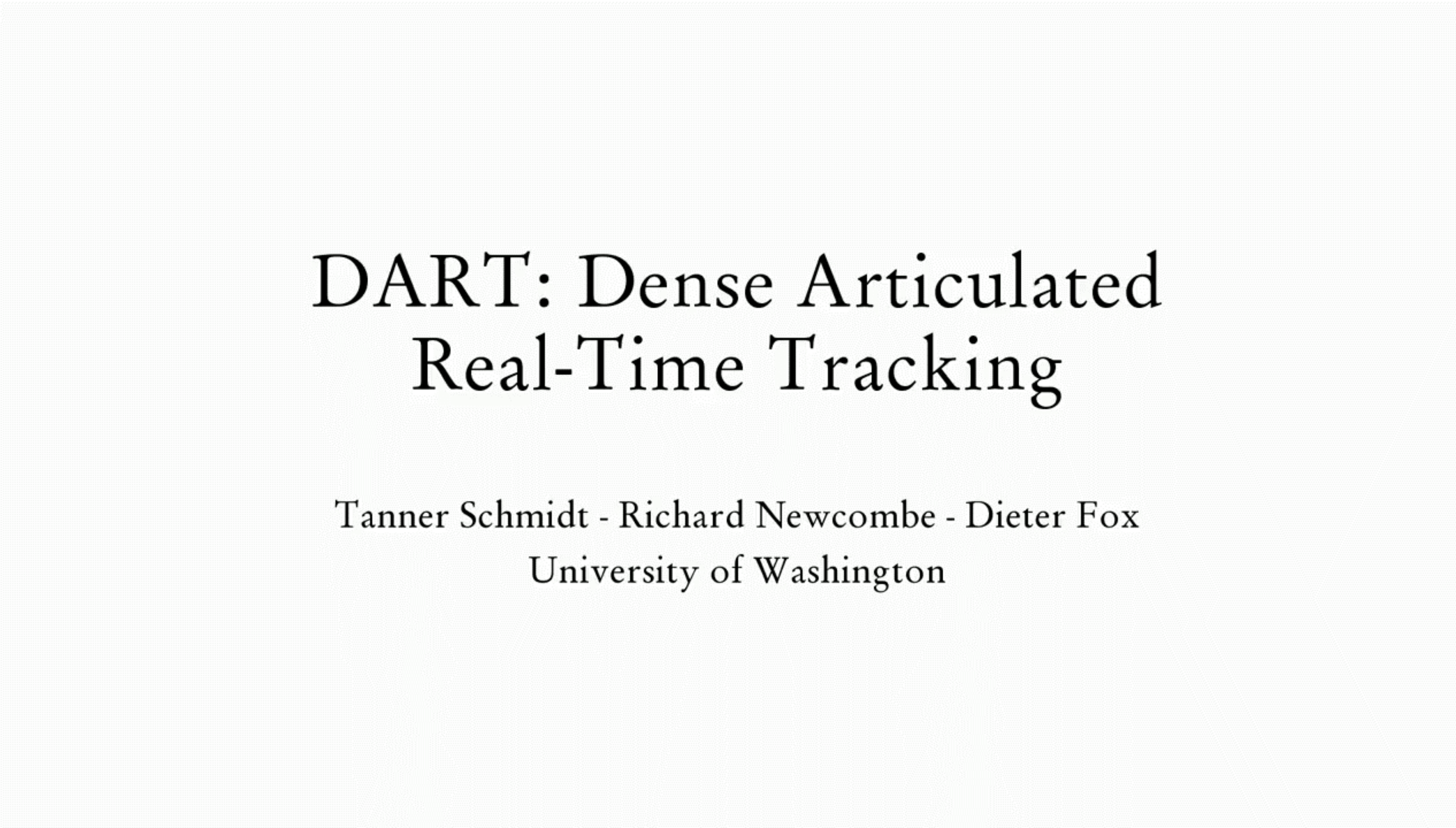
3.4 RGB-D: Applications

❖ RGB-D SLAM



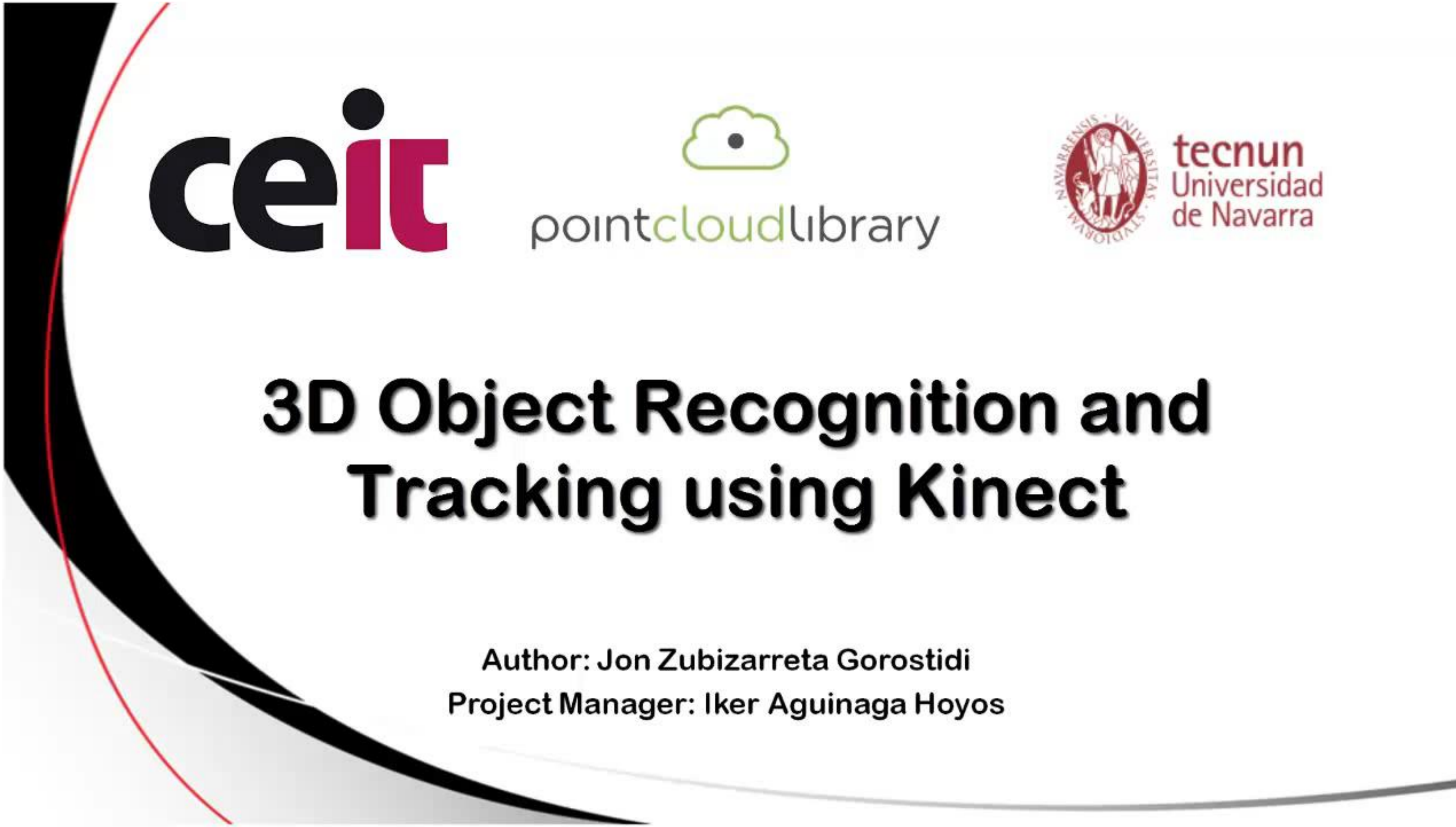
3.4 RGB-D: Applications

❖ **Real-Time Tracking**



3.4 RGB-D: Applications

- 3D Object Recognition



UTS:CAS
CENTRE FOR AUTONOMOUS SYSTEMS

41014 Sensors and Control for Mechatronic Systems

Next Lectures

Dr. **Liang Zhao**

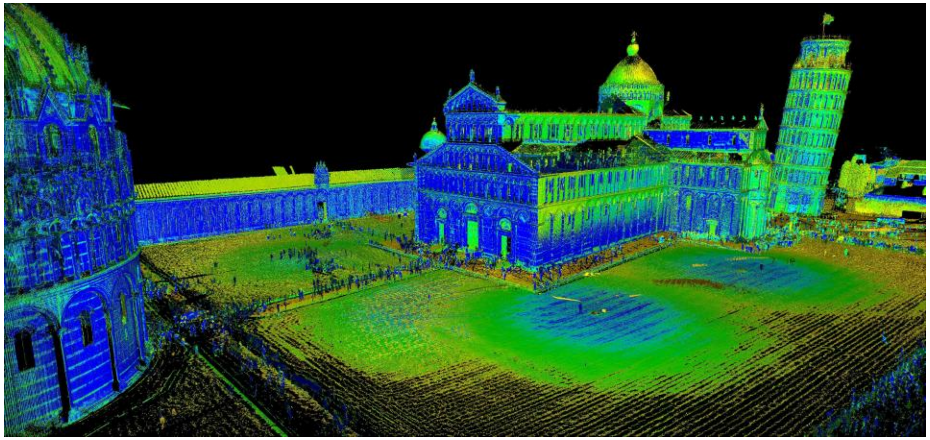
Centre for Autonomous Systems
University of Technology,
Sydney

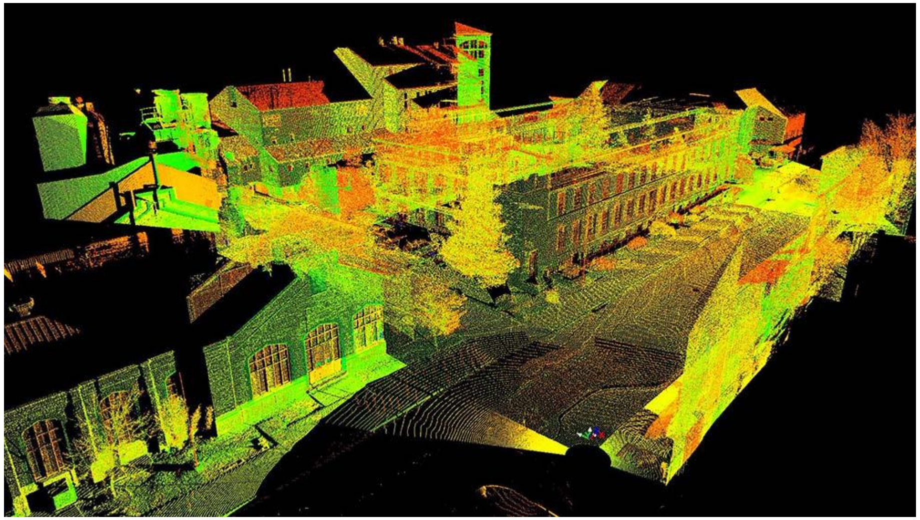
UNIVERSITY OF TECHNOLOGY SYDNEY

6. Next Lectures

- **Sensors:**
  - Cameras
  - RGB-D Cameras
  - **ToF sensors**

- **In details of**
  - Fundamental
  - Data and Processing





THANK YOU

Questions?