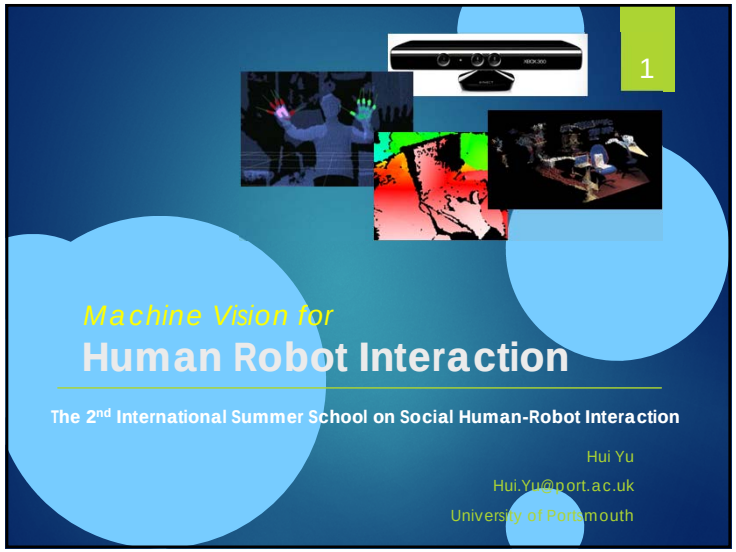
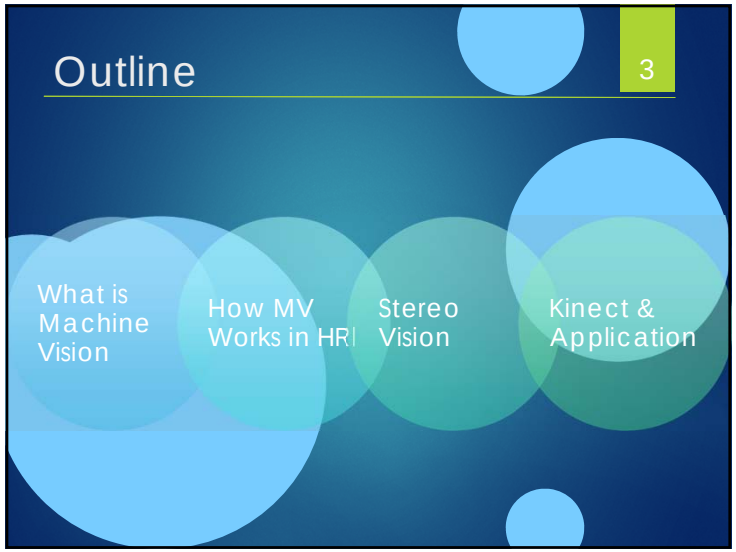
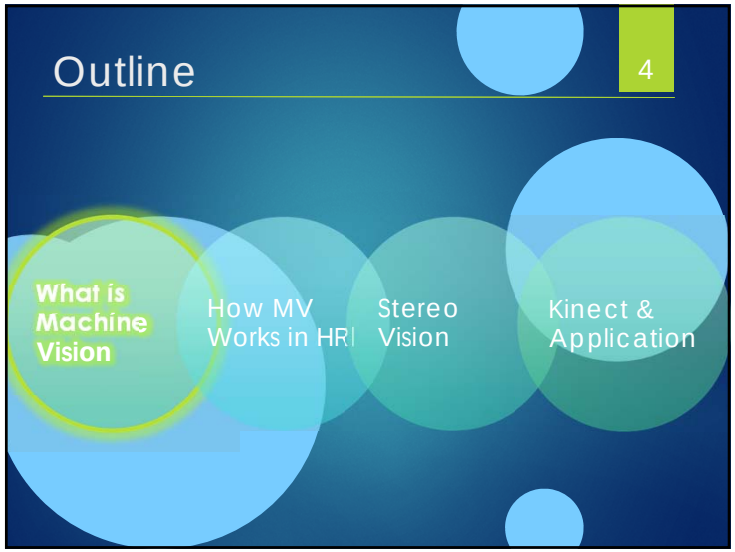

1
Machine Vision for
Human Robot Interaction
The 2nd International Summer School on Social Human-Robot Interaction
Hui Yu
Hui.Yu@port.ac.uk
University of Portsmouth


University of Portsmouth
2
Portsmouth City


Outline
3
What is Machine Vision
How MV Works in HRI
Stereo Vision
Kinect & Application


Outline
4
What is Machine Vision
How MV Works in HRI
Stereo Vision
Kinect & Application

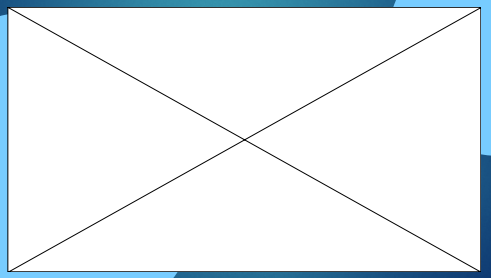
What is Machine Vision 5

Google Self-driving Car

https://www.youtube.com/watch?v=cdgQpa1pUUE

What is Machine Vision 6

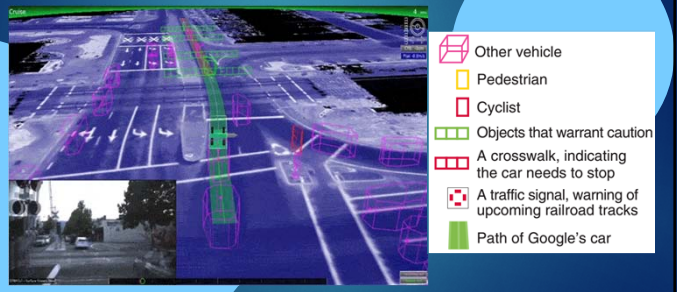
Google Self-driving Car



https://www.youtube.com/watch?v=cdgQpa1pUUE

What is Machine Vision 7

Google Self-driving Car



http://www.extremetech.com/extreme/189486-how-googles-self-driving-cars-detect-and-avoid-obstacles

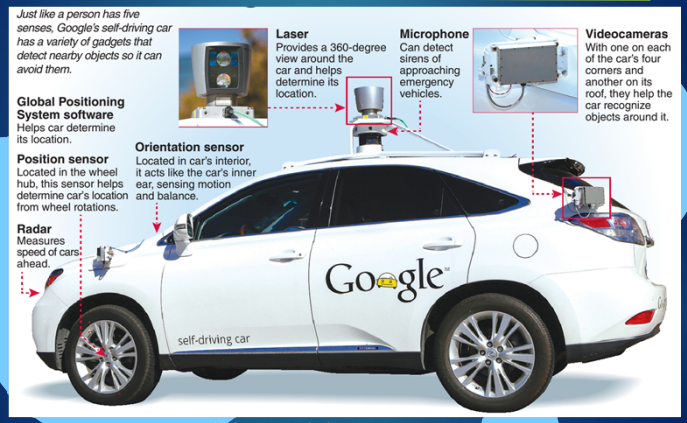
What is Machine Vision 8

Google Self-driving Car

- Detection System
  - sonar devices,
  - stereo cameras,
  - lasers
  - radar

http://www.extremetech.com/extreme/189486-how-googles-self-driving-cars-detect-and-avoid-obstacles

What is Machine Vision 9



What is Machine Vision 10

Google Self-driving Car

- Detection System
  - sonar devices, stereo cameras, Lasers and radar
- **These components have different ranges and fields of view**
- **Each serves a particular purpose.**

http://www.extremetech.com/extreme/189486-how-googles-self-driving-cars-detect-and-avoid-obstacles

What is Machine Vision 11

Google Self-driving Car

- Detection System
  - Sonar devices:
    - 6 meters
  - Stereo cameras: normal cameras in pairs with a small separation to create parallax (as human eyes)
    - 50-degree field of view
    - accurate up to about **30 meters**
  - Lasers: LIDAR system
    -highly accurate up to a range of **100 meters**
  - Radar
    - works up to 200 meters away,

http://www.extremetech.com/extreme/189486-how-googles-self-driving-cars-detect-and-avoid-obstacles

What is Machine Vision 12

Google Self-driving Car

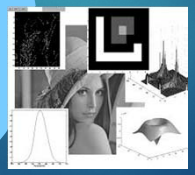
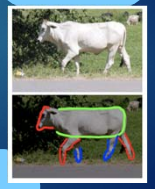
- Detection System
  - Sonar devices:
    - 6 meters
  - Stereo cameras: normal cameras in pairs with a small separation to create parallax (as human eyes)
    - 50-degree field of view
    - accurate up to about 30 meters
  - Lasers: LIDAR system
    -highly accurate up to a range of **100 meters**
  - Radar
    - works up to 200 meters away,

http://www.extremetech.com/extreme/189486-how-googles-self-driving-cars-detect-and-avoid-obstacles

What is Machine Vision 13

Introduction

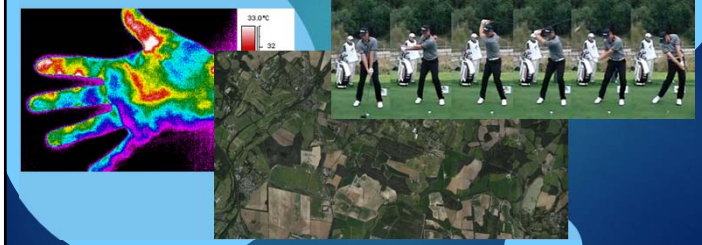
- Machine Vision is a field that includes methods for acquiring, processing, analyzing, and understanding images.
  - Image Analysis, Scene Analysis, Image Understanding, etc.
- Duplicate the abilities of Human Vision by electronically perceiving and understanding an image.

What is Machine Vision 14

Introduction

- Image data can take many forms, such as a video sequence, depth images, views from multiple cameras, infrared cameras, satellite sensors, etc.



What is Machine Vision 15

The ultimate goal of computer vision

- To model, replicate, and more importantly exceed human vision using computer software and hardware at different levels.
- It involves knowledge in computer science, electrical engineering, mathematics, physiology, biology, and co ience.

DREAM

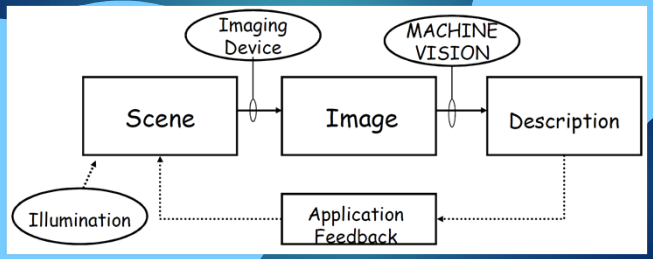
Development of Robot-Enhanced therapy for children with AutisM spectrum disorders

D R E A M

What is Machine Vision 16

Machine Vision System

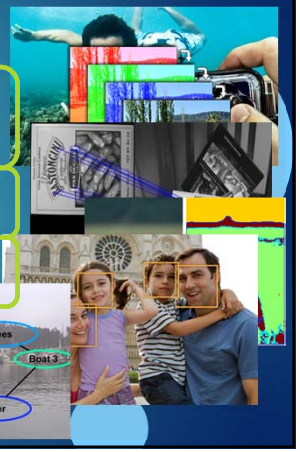
- A typical machine vision system



What is Machine Vision 17

Machine Vision Pipeline

- Image acquisition
- Multi-channel images processing
- Feature extraction and matching
- Segmentation
- Detection / Recognition
- Interpretation / Understanding
- ...



High Level Vision

What is Machine Vision 18

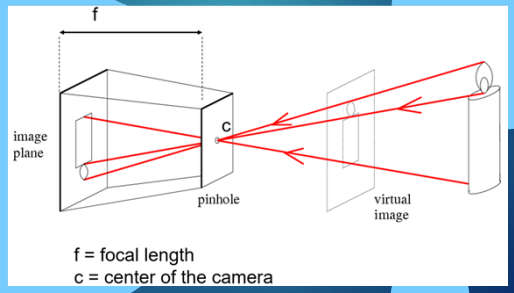
Machine Vision System

- Image acquisition
  - Mapping between image and world coordinates
    - Pinhole camera model
    - Projective geometry
      - Vanishing points and lines
    - Projection matrix

What is Machine Vision 19

Machine Vision System

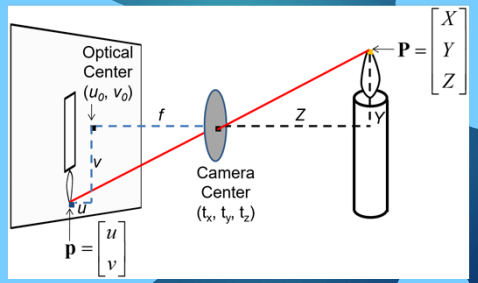
- Image acquisition
  - Pinhole camera model



What is Machine Vision 20

Machine Vision System

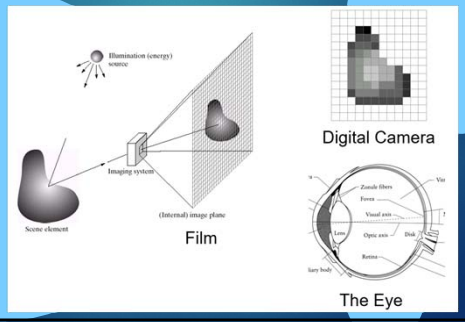
- Image acquisition
  - Projective geometry



What is Machine Vision 21

Machine Vision System

- Multi-channel images processing
  - Image Formation (Pixels)



What is Machine Vision 22

Machine Vision System

- Multi-channel images processing
  - Color Image

What is Machine Vision 23

Machine Vision System

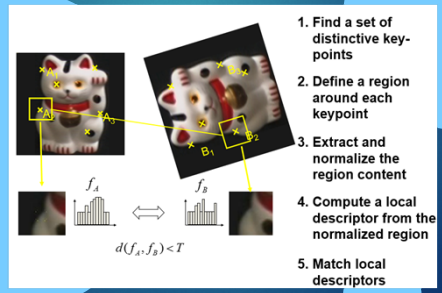
- Feature extraction and matching
  - Describe the feature

- Most features can be thought of as templates, histograms (counts), or combinations
  - Robust
  - Distinctive
  - Compact
  - Efficient
- Most available descriptors focus on edge/gradient information
  - Capture texture information
  - Color rarely used

What is Machine Vision 24

Machine Vision System

- Feature extraction and matching
  - Match the descriptors

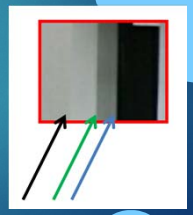


1. Find a set of distinctive key-points
2. Define a region around each keypoint
3. Extract and normalize the region content
4. Compute a local descriptor from the normalized region
5. Match local descriptors

What is Machine Vision 25

Machine Vision System

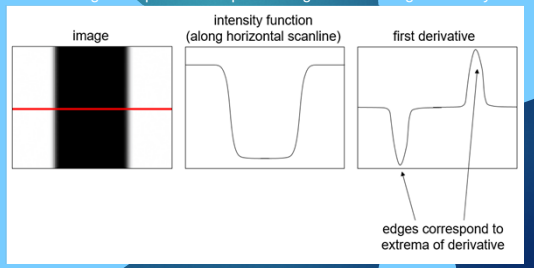
- Feature extraction and matching
  - Edge detection



What is Machine Vision 26

Machine Vision System

- Feature extraction and matching
  - Edge detection
    - An edge is a place of rapid change in the image intensity function



What is Machine Vision 27

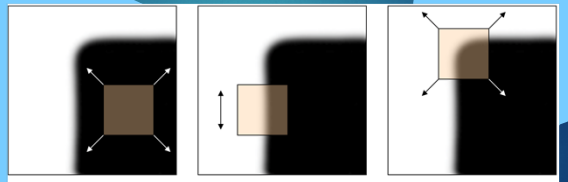
Machine Vision System

- Feature extraction and matching
  - Edge detection
    - Typical detector – Canny edge detector

What is Machine Vision 28

Machine Vision System

- Feature extraction and matching
  - Corner detection



"flat" region: no change in all directions

"edge": no change along the edge direction

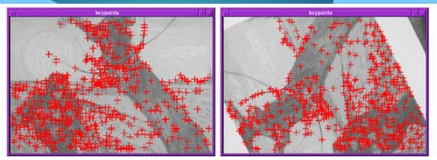
"corner": significant change in all directions

What is Machine Vision

29

Machine Vision System

- Feature extraction and matching
  - Example of Corner detection



- Key property: in the region around a corner, image gradient has two or more dominant directions
- Corners are repeatable and distinctive

What is Machine Vision

30

Machine Vision System

- Feature extraction and matching
  - Most basic corner detector – Harris corner detector
    - Compute M matrix for each image window to get their cornerness scores.
    - Find points whose surrounding window gave large corner response (f> threshold)
    - Take the points of local maxima, i.e., perform non-maximum suppression

What is Machine Vision

31

Machine Vision System

- Feature extraction and matching
  - Most basic corner detector – Harris corner detector

What is Machine Vision

32

Machine Vision System

- Feature extraction and matching
  - Other corner detector
    - DoG, SIFT, SURF, ORB
    - …

Outline

33

What is Machine Vision

How MV Works in HRI

Imaging Systems

Introducing the Kinect

How MV Works for HRI

34

Machine Vision for Human Robot Interaction

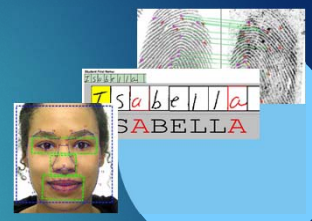
- Person identification
  - Face recognition, gender recognition, age estimation
  - Speaker identification
  - Gait recognition, clothing (recognition from a distance)
- Recognition of emotions
  - Facial expression recognition
  - Speech, body movements
- Methods for human-robot communication
  - Hand gestures and speech ("face to face" communication)
  - Body movements (e.g. waving hands) (communication from a distance)
  - Using camera motion (communication with a mobile device)
  - A talking avatar on robot's flat panel display

How MV Works for HRI

35

> Person Identification

- Authorization
  - Fingerprints verification
  - Handwrite signature
  - Facial verification
  - …



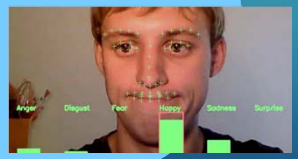
- Targeting different groups of clients / users
  - Age recognition
  - Gender recognition
  - Facial recognition
  - …

How MV Works for HRI

36

> Recognition of emotions

- Human's emotion could be a very important background context for interaction between H & R
  - Facial expression recognition
  - Speech, body movements



How MV Works for HRI 37

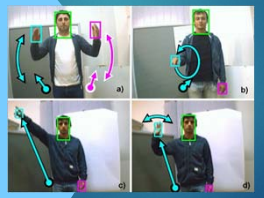
> Methods for human-robot communication

- Hand gestures and speech ("face to face" communication)
  - Gesture recognition
  - Speech recognition
- Body movements (e.g. waving hands, communication from a distance)
  - Skeleton detection
- Using camera motion (communication with a mobile device)
  - Feature matching
  - 3D reconstruction

How MV Works for HRI 38

>>> Methods for human-robot communication

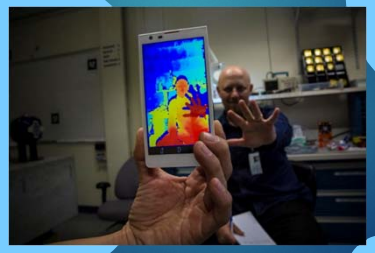
- Hand gestures and speech ("face to face" communication)
  - Gesture recognition
  - Speech recognition



How MV Works for HRI 39

>>> Methods for human-robot communication

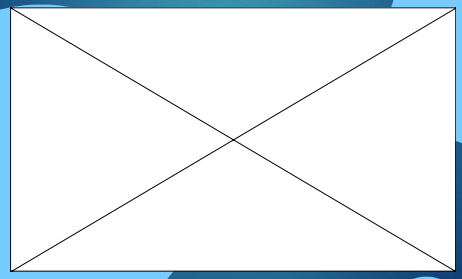
- Using camera motion (communication with a mobile device)
  - Feature matching
  - 3D reconstruction



From Google's Project Tango

Applications of HRI with MV 40

Gesture recognition for Human Robot Interaction



What is Machine Vision

41

Machine Vision System

- Stereo Vision
  - Structure and depth are inherently ambiguous from single views

Outline

42

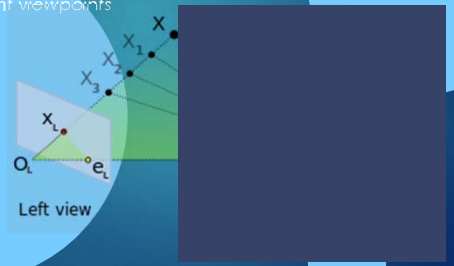
What is Machine Vision | How MV Works in HRI | **Stereo Vision** | Kinect & Application

Stereo Vision

43

Machine Vision System

- Stereo Vision
  - shape from "motion" between two views
  - infer 3d shape of scene from two (multiple) images from different viewpoints

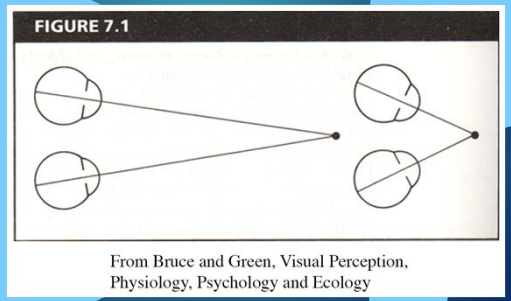


Stereo Vision

44

Machine Vision System

- Stereo Vision
  - Human stereopsis

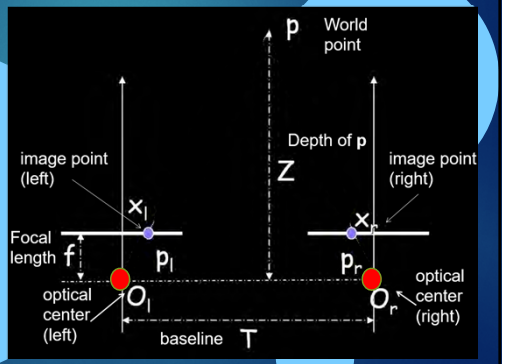


From Bruce and Green, Visual Perception, Physiology, Psychology and Ecology

Stereo Vision 45

Machine Vision System

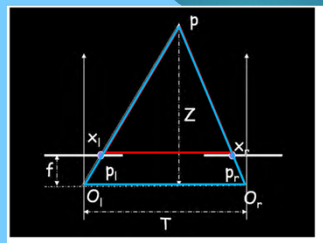
- Stereo Vision



Stereo Vision 46

Machine Vision System

- Stereo Vision
  - Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). What is expression for Z?



Similar triangles (p_l, P, p_r) and (O_l, P, O_r):

$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z}$$

$$Z = f\frac{T}{x_r - x_l}$$

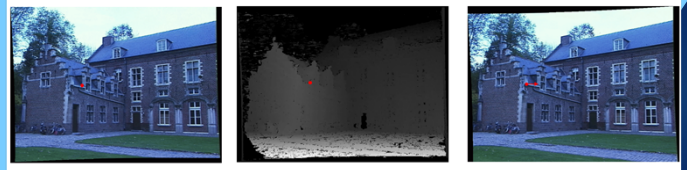
disparity

Stereo Vision 47

Machine Vision System

- Stereo Vision
  - So if we can find the corresponding points in two images, we could estimate relative depth...

image I(x,y) | Disparity map D(x,y) | image I´(x´,y´)

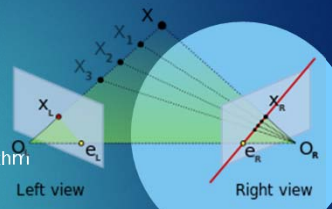


$$(x´,y´)=(x+D(x,y), y)$$

Stereo Vision 48

Machine Vision System

- Stereo Vision
- Basic stereo matching algorithm

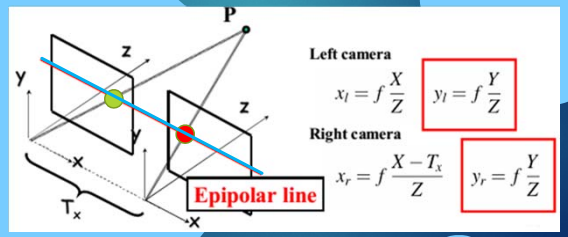


  - If necessary, rectify the two stereo images to transform epipolar lines into scanlines
  - For each pixel x in the first image
    - Find corresponding epipolar scanline in the right image
    - Examine all pixels on the scanline and pick the best match x'
    - Compute disparity x-x' and set depth(x) = fB/(x-x')

Stereo Vision 49

Machine Vision System

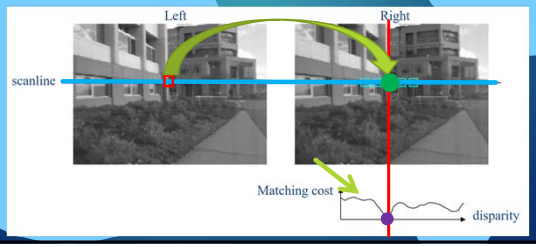
- Stereo Vision
- Basic stereo matching algorithm



Stereo Vision 50

Machine Vision System

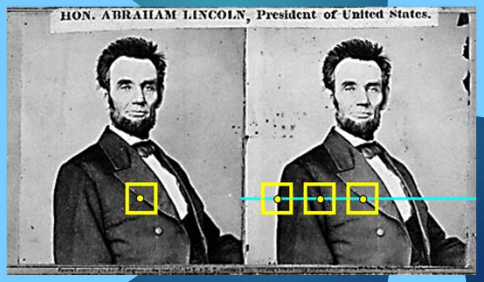
- Stereo Vision
  - Basic stereo matching algorithm
    - Slide a window along the right scanline and compare contents of that window with the reference window in the left image
    - Matching cost: SSD or normalized correlation



Stereo Vision 51

Machine Vision System

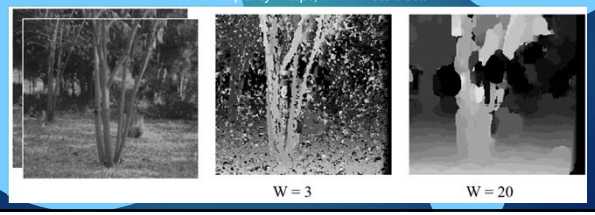
- Stereo Vision
  - Basic stereo matching algorithm



Stereo Vision 52

Machine Vision System

- Stereo Vision
  - Effect of window size
    - Smaller window
      - \+ More detail; – More noise
    - Larger window
      - \+ Smoother disparity maps; – Less detail



Outline

53

What is Machine Vision

How MV Works in HRI

Stereo Vision

Kinect& Application

54

Introducing to the device of

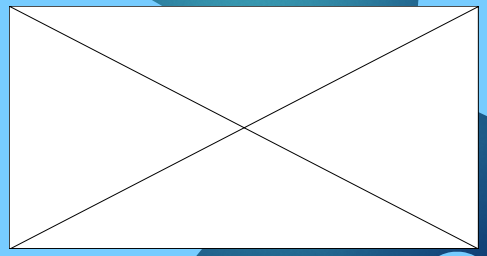
Kinect & Application

Applications of HRI with MV

55

Kinect based human action recognition

- Real Time Action Recognition based on Space-time Occupancy Patterns for Human Robot Interaction



Outline

56

1. What is the Kinect
2. How does it work
3. How the Kinect compares to other sensors
4. How can we use the Kinect
5. Applications of the Kinect

Outline

57

- What is the Kinect
- How does it work
- How the Kinect compares to other sensors
- How can we use the Kinect
- Applications of the Kinect

What is the Kinect

58

First look of the Kinect

- Kinect was launched in North America on 4 November 2010
- It is used for Xbox 360 console and Windows PCs
- Enable interaction: with game/application
  - Without the need to touch a controller; use gestures and spoken

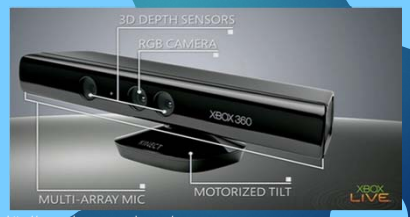
http://www.ehomeupgrade.com/wp-content/uploads/xbox_360_kinect_diagram.jpg

What is the Kinect

59

First look of the Kinect

- Kinect is a depth camera
- It gives you the depth values associated with every pixel
- It uses structured infrared light to determine depth values.



http://www.ehomeupgrade.com/wp-content/uploads/xbox_360_kinect_diagram.jpg

What is the Kinect

60

First look of the Kinect

- Software is developed by **Rare**
- Camera technology is developed by Israeli **PrimeSense**

- **Kinect for Windows**
  - Launched In February 2012
  - Enables the depth camera to see objects as close as 40 cm
  - Skeletal tracking enhancement
  - …

What is the Kinect

61

First look of the Kinect

- Work Region
  - First Region: a high accurate depth surface for near objects aprox. (0.8 – 1.2 m)
  - Second Region: medium accurate depth surface aprox. (1.2 – 2.0 m).
  - Third Region: Allows to obtain a low accurate depth surface in far objects aprox. (2.0 – 3.5 m).

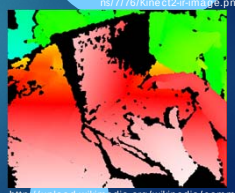
What is the Kinect

62

First look of the Kinect

- Gesture and Facial Recognition
  - RGB Camera
  - Depth Camera
    - Infrared Laser Projector
    - Monochrome CMOS Sensor
- Voice Recognition
  - Multi-Array Microphone

http://upload.wikimedia.org/wikipedia/commons/7/76/Kinect2-ir-image.png



http://upload.wikimedia.org/wikipedia/commons/9/90/Kinect2-deepmap.png

What is the Kinect

63

Why was the Kinect Made

- Make UI More Intuitive to Human Motion
- Release a Competitor to Nintendo WII
- Led to More Affordable 3D Imaging and Motion Detection

Outline

64

1. What is the Kinect
2. How does it work
3. How the Kinect compares to other sensors
4. How can we use the Kinect
5. Applications of the Kinect

What is the Kinect

65

First look of the Kinect

- The Kinect uses an infrared projector and sensor;
- The technique of analysing a known pattern is called *structured light*

Note: it does not use its RGB Camera for depth computatio

Picture from [1]

What is the Kinect

66

First look of the Kinect

- Principles of Kinect (Primesensor)
  - Time Multiplexing
  - Direct coding
  - Spatial Neighbourhood

This coding has to be unique per position in order to recognize each point in the patt

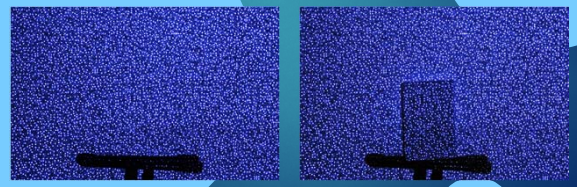
http://www.ehomeupgrade.com/wp content/uploads/xbox_360_kinect_diagram.jpg

How does it work

67

Depth Sensor

- IR Projector emits predefined Dotted Pattern
- Lateral Shift between projector and sensor
  - Shift in pattern dots
- Shift in Dots determines Depth of Region



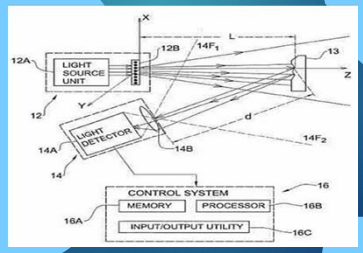
http://www.futurepicture.org/?p=116

What is the Kinect

68

First look of the Kinect

- How Kinect works?
  - Projects a known pattern (Speckles) in Near-Infrared light.
  - CMOS IR camera observes the scene.

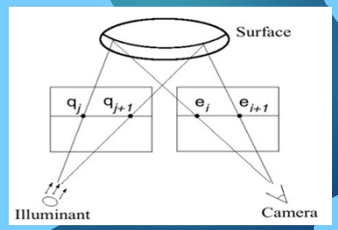


What is the Kinect

69

First look of the Kinect

- How calculate the depth data?
- Triangulation of each speckle between a virtual image (pattern) and observed pattern.
- Each point has its correspondence speckle.

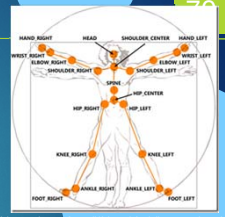


How does it work

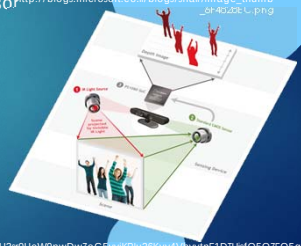
70

Nature of the Functions

- Infrared Projection
  - Dot Pattern
- Motion Detection
  - 30 Hz Output from CMOS Sensor
  - 40 Tracking Points
    - 20 Points per player
- Depth Mapping
  - Object Triangulation
- Device Reaction



http://blogs.microsoft.co.il/blogs/shair/image_thumb_6F4B2BEC.png



http://api.ning.com/files/46-6*7H3sr0HeW9nwDwZoG5vvjKRlu26Kvy4VhyytnE1DTUjsfQ5OZFQ5g8LFS-ll5DwjPVHn5NMQl*UgohBCpXU-W9b82XWPpmeSensorDepth_Diagram.gif

How does it work

71

Depth Sensor

- Field of View
  - 57 Degree Horizontal, 43 Vertical
  - Movable by additional 27 Degrees Up or Down
- IR Projector
- Received by Monochrome CMOS Sensor
- Video Output
  - 9 – 30 Hz
  - 640 x 480px
- Additional Processing
  - IR Filter at Projector Wavelength

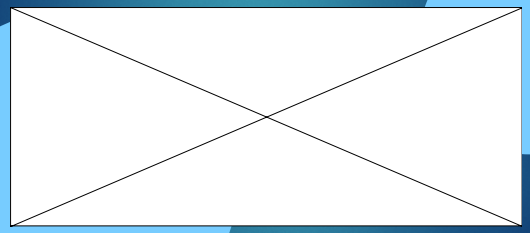
http://www.kinecttr.com/wp-content/uploads/2011/05/kinect_hack.jpg

How does it work

72

Depth Sensor

- IR Dotted Pattern



https://www.youtube.com/watch?v=dTKlNGSH9Po

Outline 73

- What is the Kinect
- How does it work
- How the Kinect compares to other sensors
- How can we use the Kinect
- Applications of the Kinect

Compares to other sensors 74

ASUS Xtion / PrimeSense Carmine

Sensors	Pros	Cons
 Microsoft Kinect	➢ High quality of device drivers ➢ Stable work ➢ Has motor that can be controlled remotely which makes device positioning more convenient	➢ Bigger size ➢ Higher weight ➢ Require ACDC power supply ➢ Higher interference with another Kinect sensor in "Dual depth sensor" configuration ➢ Lower RGB image quality in comparison with MS Kinect
 PrimeSense Carmine 1.08	➢ More compact ➢ Lighter weight ➢ Does not require power supply except USB ➢ Lower interference with another ASUS Xtion / PrimeSense Carmine sensor in "Dual depth sensor" configuration ➢ Better RGB image quality	➢ Less popular device ➢ Lower drivers quality ➢ Does not work with some USB controllers (especially USB 3.0) ➢ No motor, allow only manual positioning

ASUS Xtion / PrimeSense Carmine

Compares to other sensors 75

ASUS Xtion / PrimeSense Carmine

Sensors

Microsoft Kinect

Kinect is more powerful

PrimeSense Carmine 1.08

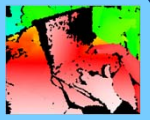
ASUS Xtion / PrimeSense Carmine

Compares to other sensors 76

Stereo Cameras

Kinect

- Pros:
  - Higher accuracy;
  - More reliable;
  - Less calibration process;
  - More features related to recognitions
- Cons:
  - More expensive;



Stereo Camera

- Pros:
  - Cheaper;
  - High resolution capability;
  - Flexible sensor shape
- Cons:
  - Need complicated calibration process;
  - Rely on illumination conditions

Compares to other sensors 77

Time-of-Flight cameras

Kinect
- Pros:
  - Less expensive
  - High precision
- Cons:
  - Short range indoor measurement

Time-of-Flight Cameras
- Pros:
  - Long range outdoor measurement;
  - Less shadow effect
- Cons:
  - Much more expensive
  - Depth Info is much nosier
  - Hard to reach higher resolution

Outline 78

1. What is the Kinect
2. How does it work
3. How the Kinect compares to other sensors
4. How can we use the Kinect
5. Applications of the Kinect

How to use the Kinect 79

Microsoft Kinect SDK

- Kinect Software Development Kit (SDK)
  - Free Download available for Windows 7 and 8
  - Official SDK
  - Http://www.microsoft.com/enus/kinectforwindows/develop/overview.aspx
- Require Windows OS Capable Processor
- Very well documented and great online community support

How to use the Kinect 80

Other Kinect SDK

- OpenNI
  - Voice and voice command recognition
  - Hand gestures
  - Body Motion Tracking
  - Free to use, Well documented, widely used.
  - More flexible, and more friendly interface for coding.
  - Compatible with other Depth Sensor, which make it easier for code migration for different sensor.
  - http://structure.io/openni
- Other SDKs, including OpenKinect, have been ported to Linux and Mac

Outline

81

- What is the Kinect
- How does it work
- How the Kinect compares to other sensors
- How can we use the Kinect
- Applications of the Kinect

Applications of the Kinect

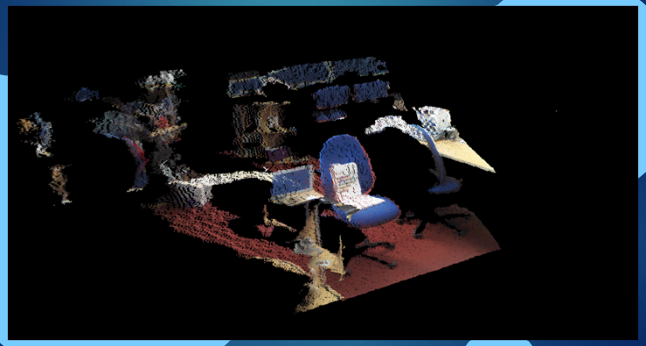
82

I. 3D Scene Mapping with a Single Shot

Applications of the Kinect

83

I. 3D Scene Mapping with a Single Shot



Applications of the Kinect

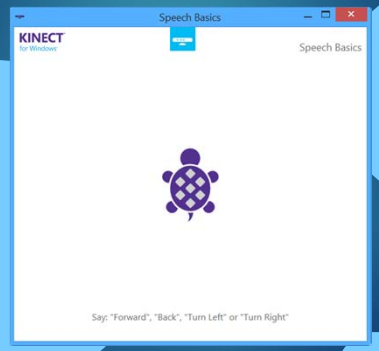
84

II. Hand Signals and Motion Recognition

http://www.popsci.com/files/imagecache/article_image_large/articles/Picture%202_54.png

Applications of the Kinect 85

III. Voice Recognition



From Kinect Developer Toolkit

Applications of the Kinect 86

IV. Object Tracking

http://geektech.ie/wp-content/uploads/2011/04/eye-object-tracking.jpg

http://hackadaycom.files.wordpress.com/2011/04/tld_object_tracking.jpg?w=470&h=313

Applications of the Kinect 87

Human Machine Interaction in Games

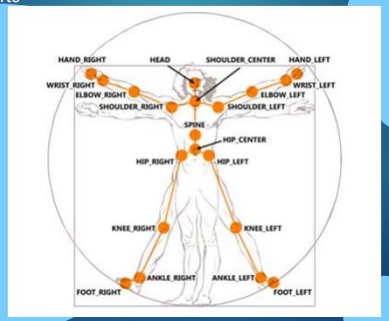
- Kinect was first introduced with Xbox 360, a game console.

http://agrlv.com/wp-content/uploads/2013/09/XBox-Kinect-with-50-in-Screen-with-Sports-software.jpg

Applications of the Kinect 88

II. Skeleton Detection

- Using skeletonstream, Kinect SDK provides us with 20 joints



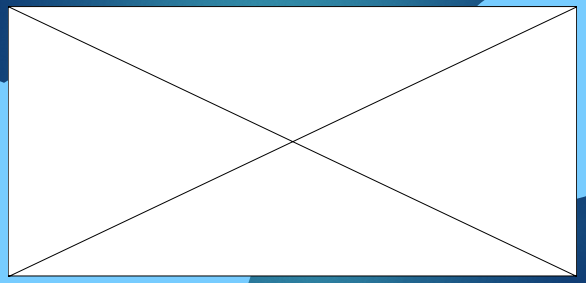
Applications of the Kinect 89

Human Machine Interaction in Games

- 1. Game experience using Skeleton Recognition.
  - Body gesture controls features in games

Applications of the Kinect 90

Applications



From [2]

Applications of the Kinect 91

V. Applications using combinations of Kinect' s features

- Inferring body position is a two-stage process:
  - First , compute a depth map (using structured light)
  - Second, infer body position (using machine learning)

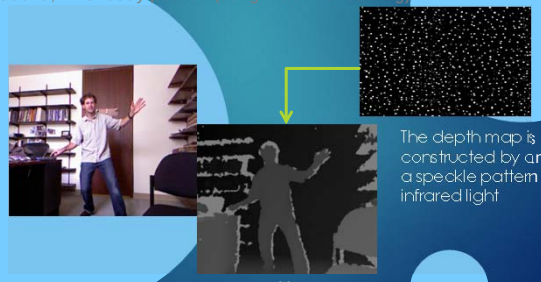
From [2]

Applications of the Kinect 92

V. Applications using combinations of Kinect' s features

Inferring body position is a two-stage process:

- First , compute a depth map (using structured light)
- Second, infer body position (using machine learning)



The depth map is constructed by analysing a speckle pattern of infrared light

From [2]

Applications of the Kinect 93

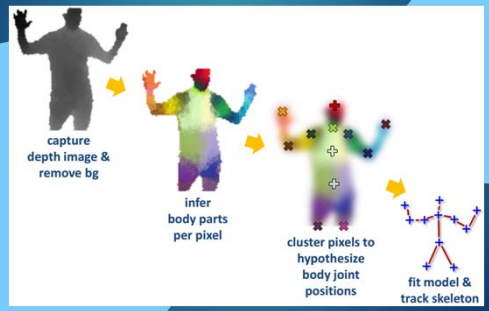
V. Applications using combinations of Kinect' s features

- Inferring body position is a two-stage process:
  - First , compute a depth map (using structured light)
  - Second, infer body position (using machine learning)

From [2]

Applications of the Kinect 94

V. Applications using combinations of Kinect' s features



Slides from Shotton et al. CVPR2011 presentation

Applications of the Kinect 95

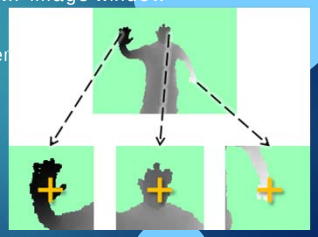
V. Applications using combinations of Kinect' s features

- Infer Body part
  - Compute P(Cx | Wx)

- where x-pixel, Cx – body part, Wx- image window

  - Discrimination

Use training data to learn classifier



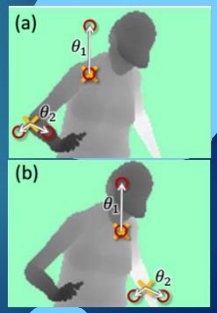
Slides from Shotton et al. CVPR2011 presentation

Applications of the Kinect 96

V. Applications using combinations of Kinect' s features

- Depth Image features
  - Depth comparison
  - In combination in a decision forest

which part of the body the pixel belongs to



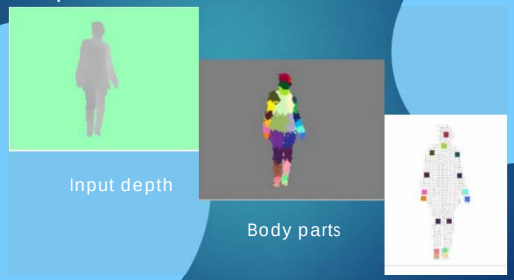
Slides from Shotton et al. CVPR2011 presentation

Applications of the Kinect

97

V. Applications using combinations of Kinect' s features

- Example



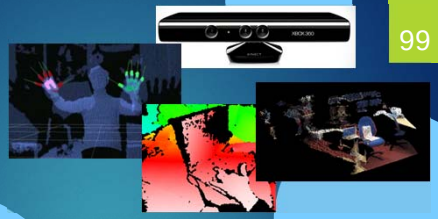
Slides from Shotton et al. CVPR2011 presentation

Applications of the Kinect

98

V. Applications using combinations of Kinect' s features

From [2]



99

The **2nd International Summer School on Social Human-Robot Interaction**

Questions

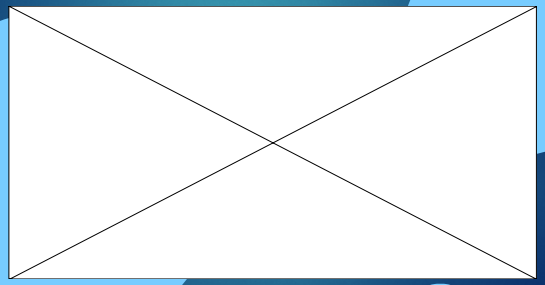
Hui Yu
Hui.Yu@port.ac.uk
University of Portsmouth

Applications of the Kinect

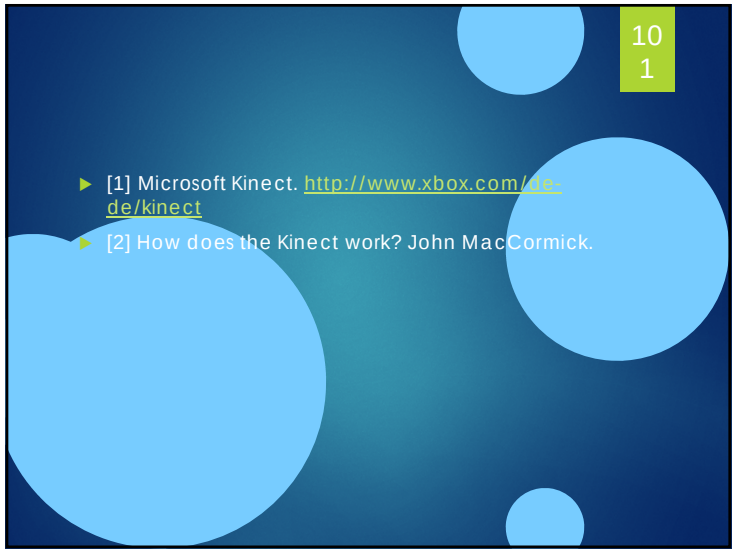
100

II. 3D Reconstruction

- Kinect human-following robot



https://www.youtube.com/watch?v=I_fre0E_J8s


10
1
[1] Microsoft Kinect. http://www.xbox.com/de-de/kinect
[2] How does the Kinect work? John MacCormick.