

Introduction: Sensing Visual Data in ROS

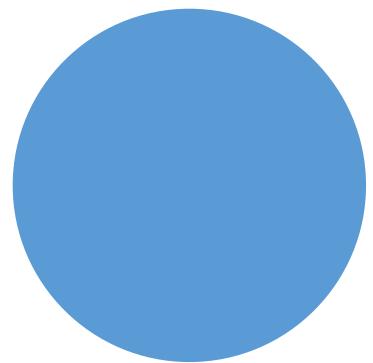
github.com/Tobias-Fischer/HCR-Vision-Tutorial

Human-Centred Robotics

Dr Tobias Fischer

2019-11-07





Computer Vision |
The very basics |

Computer Vision: The very basics

First encounter

088 084 161 169 152 053 123 221 119 121 144 138 152 164 158 103 215 192 098 232 183 161 178 177 071 072 151 188 175 065 139 237 118 084 116 089 168 119 118 085 218 210 114 246 168 143 169 168 025 036 106 153 166 055 123 226 098 038 044 027 039 042 046 049 215 209 107 232 128 084 116 124
105 059 141 117 086 037 086 288 144 179 179 153 187 183 185 113 228 184 078 135 182 162 157 168 088 057 146 132 088 041 091 220 125 138 138 147 149 145 094 219 202 093 154 184 192 148 156 037 029 117 112 065 027 066 187 076 059 057 052 055 059 062 041 195 194 088 145 188 081 094 108
078 000 161 076 079 193 189 149 152 178 166 182 181 166 192 121 176 162 135 097 124 057 139 168 065 069 103 075 084 189 114 152 118 126 138 131 131 121 143 091 166 180 153 188 133 064 123 154 035 008 082 065 078 092 093 115 051 048 045 049 043 039 052 032 115 164 156 185 128 048 088 108
117 039 126 129 091 078 102 147 117 122 179 189 171 158 184 143 188 181 086 128 185 119 111 104 089 032 130 143 108 079 114 071 074 124 117 089 138 165 157 108 045 011 107 133 102 075 116 150 062 013 046 056 034 024 058 030 153 161 103 143 198 094 056 064
090 013 060 042 031 058 028 183 156 088 156 188 182 186 188 143 203 063 051 045 061 047 122 069 068 089 054 046 026 056 025 116 167 086 189 145 138 145 151 112 222 078 058 038 069 052 106 064 014 004 038 044 085 049 019 116 151 015 030 068 056 061 063 044 213 068 049 017 066 039 061 038
090 073 105 108 055 094 068 018 088 063 112 154 179 163 144 131 101 081 121 083 148 105 106 121 056 049 079 116 048 077 053 019 082 038 086 111 133 121 103 098 101 078 108 074 157 097 084 109 014 014 031 109 034 027 021 016 066 007 010 034 047 037 028 026 077 039 051 052 154 056 057 062
081 101 132 098 086 071 048 089 028 067 104 114 135 128 126 124 050 017 081 119 093 127 151 118 081 078 101 085 099 071 041 089 012 033 088 078 084 075 083 085 030 018 088 134 093 107 131 102 082 028 041 051 102 065 034 088 000 003 014 016 028 012 019 016 088 009 078 147 074 054 073 068
099 005 088 122 085 123 097 028 072 095 084 102 168 095 109 131 107 066 162 087 093 174 131 092 087 085 078 104 078 117 084 013 088 040 063 069 057 078 095 070 056 102 088 083 158 122 089 042 047 043 055 045 048 044 088 007 012 068 012 008 019 024 015 022 133 071 048 111 078 063
122 092 152 124 125 123 083 071 109 187 083 105 162 096 098 148 153 102 120 152 157 198 117 106 122 087 138 106 108 098 049 039 087 078 052 089 065 065 065 111 117 063 083 134 138 181 117 108 085 058 092 055 044 039 008 008 018 023 011 015 013 012 035 044 011 025 074 088 125 088 084
157 102 087 117 103 089 128 166 153 110 101 090 083 090 112 135 152 168 127 129 158 138 153 146 164 111 085 106 085 078 088 131 120 083 069 059 058 083 078 108 127 131 091 102 141 124 162 155 139 099 058 066 038 018 037 075 054 029 017 012 014 016 019 038 059 061 037 044 099 081 148 128
114 181 182 128 082 110 163 147 125 196 073 087 118 083 126 153 179 173 112 105 116 185 178 121 208 112 157 118 057 098 141 120 086 073 048 056 088 057 094 126 159 161 103 085 113 199 181 199 178 094 126 073 022 054 088 056 038 017 097 011 023 014 028 061 101 118 082 055 094 188 162
019 038 094 168 118 088 059 106 122 077 056 029 039 077 051 084 114 168 116 059 108 206 126 217 024 046 100 204 138 089 052 082 109 059 038 013 016 045 025 056 099 151 107 059 113 222 143 228 015 033 085 108 103 065 030 005 019 067 001 063 088 002 009 042 100 063 033 098 207 131 213
053 048 111 144 109 141 124 066 142 042 031 050 036 063 104 112 075 141 089 086 144 077 085 172 057 058 117 159 118 151 131 081 084 142 042 017 025 015 044 074 083 071 149 089 158 056 084 181 043 046 089 140 102 138 112 048 115 038 004 011 068 009 024 041 045 128 048 078 139 072 075 166
088 152 119 081 034 015 042 044 145 091 140 234 176 164 204 241 142 169 098 036 034 074 131 102 091 160 139 091 081 038 018 045 044 151 099 097 154 113 107 145 199 141 184 103 032 037 082 142 110 078 139 109 071 027 012 035 025 133 088 077 139 095 085 123 161 129 162 082 024 027 062 117 088
081 108 154 083 004 051 086 085 092 170 087 178 258 248 253 168 188 148 082 114 058 107 155 108 081 108 168 093 085 056 078 072 089 188 053 117 183 167 187 132 195 158 088 127 085 116 171 182 077 087 150 085 004 044 056 082 087 175 042 188 168 158 185 125 198 144 088 119 050 101 147 072
182 110 113 102 012 015 051 094 154 185 139 035 166 172 084 187 132 107 084 187 132 086 116 155 204 122 122 110 014 015 055 122 084 145 202 111 071 082 208 201 105 072 039 118 123 162 184 103 108 085 010 016 042 092 160 089 101 017 012 014 016 029 102 101 124
155 142 157 124 034 043 127 133 121 132 149 187 084 088 083 143 158 131 129 128 089 124 074 175 169 155 182 136 033 042 139 151 137 148 164 184 111 046 047 077 158 171 151 147 143 133 082 142 082 184 147 133 132 112 024 031 127 144 125 153 153 183 048 041 071 153 168 146 125 082 102 188 159
103 166 141 150 054 049 050 103 133 148 147 151 039 015 156 166 158 156 111 062 093 149 181 192 109 175 150 161 061 081 049 055 114 150 167 163 168 042 012 016 167 213 185 173 133 071 102 158 198 208 099 147 122 134 047 037 041 093 138 159 157 162 036 086 161 220 184 166 124 063 099 122 139 172
122 082 077 128 071 045 047 092 109 116 168 159 057 052 123 142 135 111 082 086 127 171 216 193 128 088 084 138 088 044 058 089 107 111 178 133 048 045 122 138 134 117 085 078 148 183 222 208 095 059 078 101 068 035 033 053 076 092 158 128 048 037 111 127 106 089 073 066 119 141 159 153
137 128 088 098 102 040 031 141 111 122 227 226 142 166 206 132 133 069 034 104 174 166 184 146 137 093 092 109 041 027 128 081 086 207 115 173 119 100 086 038 115 174 167 176 178 109 104 087 063 033 029 015 089 037 072 186 233 115 149 173 090 038 027 023 092 132 144 127 118
102 079 086 085 037 003 123 124 087 088 083 086 089 055 086 089 041 088 083 084 089 088 023 023 027 038 101 059 061 049 082 029 009 104 044 027 027 085 022 021 019 026 076 026 031 119 113 129 112
141 028 033 067 107 029 000 142 131 090 077 069 083 073 065 063 139 012 074 131 128 202 161 176 157 034 030 068 119 032 009 123 081 038 033 029 032 017 069 020 121 011 077 141 135 210 162 150 136 024 013 044 093 024 000 075 059 038 016 011 017 001 007 018 076 005 061 113 107 173 112 085
167 110 038 029 050 021 012 038 112 133 166 119 132 131 121 045 021 022 072 151 204 154 118 187 182 124 038 018 055 028 089 033 087 087 058 069 081 068 058 029 019 028 077 162 222 168 108 158 159 108 028 012 044 015 008 025 073 068 042 051 074 062 068 033 015 014 058 139 197 138 058 081
152 159 108 022 066 012 033 005 012 121 154 145 168 168 078 011 042 052 114 206 130 126 161 188 167 177 120 023 001 018 031 008 063 077 081 089 091 049 017 043 051 128 214 144 129 136 155 139 156 101 015 001 008 024 088 010 071 082 087 103 111 061 022 048 048 109 204 124 093 069 081
150 147 152 107 023 007 049 042 079 095 169 283 217 211 063 087 087 106 045 078 070 081 171 176 178 163 184 165 116 027 007 048 039 082 047 113 138 135 137 044 088 109 045 088 089 136 178 146 134 141 143 091 017 013 047 038 078 054 136 168 171 167 053 087 107 043 070 075 109 127 074 074
152 144 143 133 094 029 028 062 233 215 243 254 251 253 211 244 116 022 021 136 151 164 175 123 178 157 153 144 102 032 028 052 229 179 206 248 240 234 182 240 119 025 024 145 162 153 142 165 147 131 127 117 078 026 025 041 191 153 235 252 254 241 150 206 110 025 014 026 126 132 097 078 056
152 144 133 122 116 048 022 035 214 228 213 220 216 212 225 208 037 023 123 159 167 188 091 068 170 158 166 133 128 055 028 026 185 198 202 263 191 038 022 157 177 123 087 058 154 142 124 166 094 038 023 013 121 149 201 219 216 207 158 134 029 022 125 162 138 119 036 038
082 146 138 126 117 066 016 093 026 029 056 053 063 075 066 037 023 061 117 153 231 125 066 176 088 163 157 146 128 076 017 003 018 046 055 045 057 088 066 038 025 066 139 165 236 121 069 141 073 147 140 118 096 057 018 004 010 049 063 045 058 089 069 037 024 062 125 146 198 089 030 072
089 086 136 133 130 093 054 071 065 085 076 086 088 084 102 112 130 141 167 241 184 166 178 197 012 084 154 149 142 104 068 072 065 088 076 089 083 108 119 143 156 182 250 184 092 135 150 066 081 136 126 119 087 043 053 047 065 052 058 063 058 081 085 119 139 171 235 170 047 058 071
104 024 064 115 110 127 139 158 145 100 077 069 094 082 133 174 190 174 154 117 099 182 184 194 123 031 073 130 148 139 146 152 148 106 079 091 090 100 082 142 186 205 190 102 123 098 144 137 151 105 020 059 116 128 119 121 110 105 079 055 068 080 083 111 156 180 179 148 116 072 071 055 070
150 082 089 031 101 124 144 150 168 141 095 117 119 092 151 209 135 064 093 080 132 204 193 192 169 086 082 038 117 138 154 153 162 151 100 121 093 168 231 151 074 004 002 116 113 146 158 146 074 000 031 099 117 137 115 115 132 080 166 118 071 137 210 139 064 089 000 067 088 063 087

What do you see ?

Slide: Vittorio Ferrari

Computer Vision: The very basics

First encounter

Computer vision is the field of study that aims at enabling computers to interpret and understand the visual world. It is a multidisciplinary field that involves computer science, mathematics, physics, and engineering. Computer vision has many applications, such as image processing, object recognition, scene understanding, and robotics.

The first step in computer vision is to capture an image. This can be done using a camera or a scanner. Once an image is captured, it needs to be processed. This involves several steps:

- Image acquisition: The image is captured by a camera or a scanner.
- Image pre-processing: The image is converted into a digital format and noise is removed.
- Feature extraction: The image is analyzed to extract features, such as edges, corners, and shapes.
- Object detection: The image is analyzed to detect objects, such as cars, people, and buildings.
- Scene understanding: The image is analyzed to understand the scene, such as the location, the time, and the weather.

Computer vision is a complex field that requires a deep understanding of mathematics, physics, and computer science. It is a rapidly growing field with many exciting applications.

And now?

Computer Vision: The very basics

First encounter

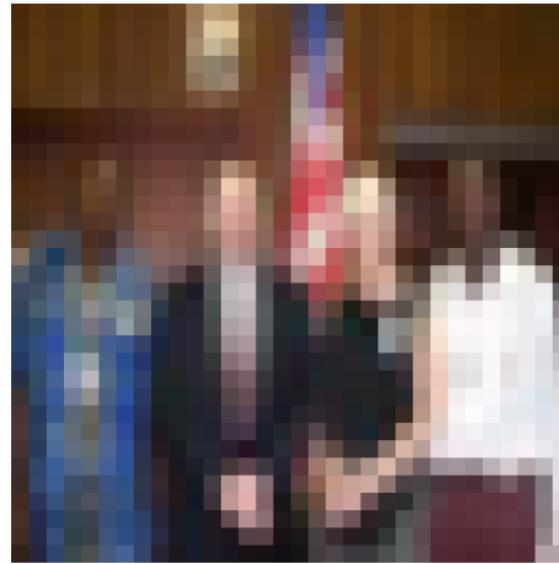


Rapid visual recognition is important !

Computer Vision: The very basics

The goal of computer vision

To extract “meaning” from pixels



Humans are remarkably good at this...

Source: “80 million tiny images” by Torralba et al.

Computer Vision: The very basics

Vision as a source of semantic information



slide credit: Fei-Fei, Fergus & Torralba

Computer Vision: The very basics

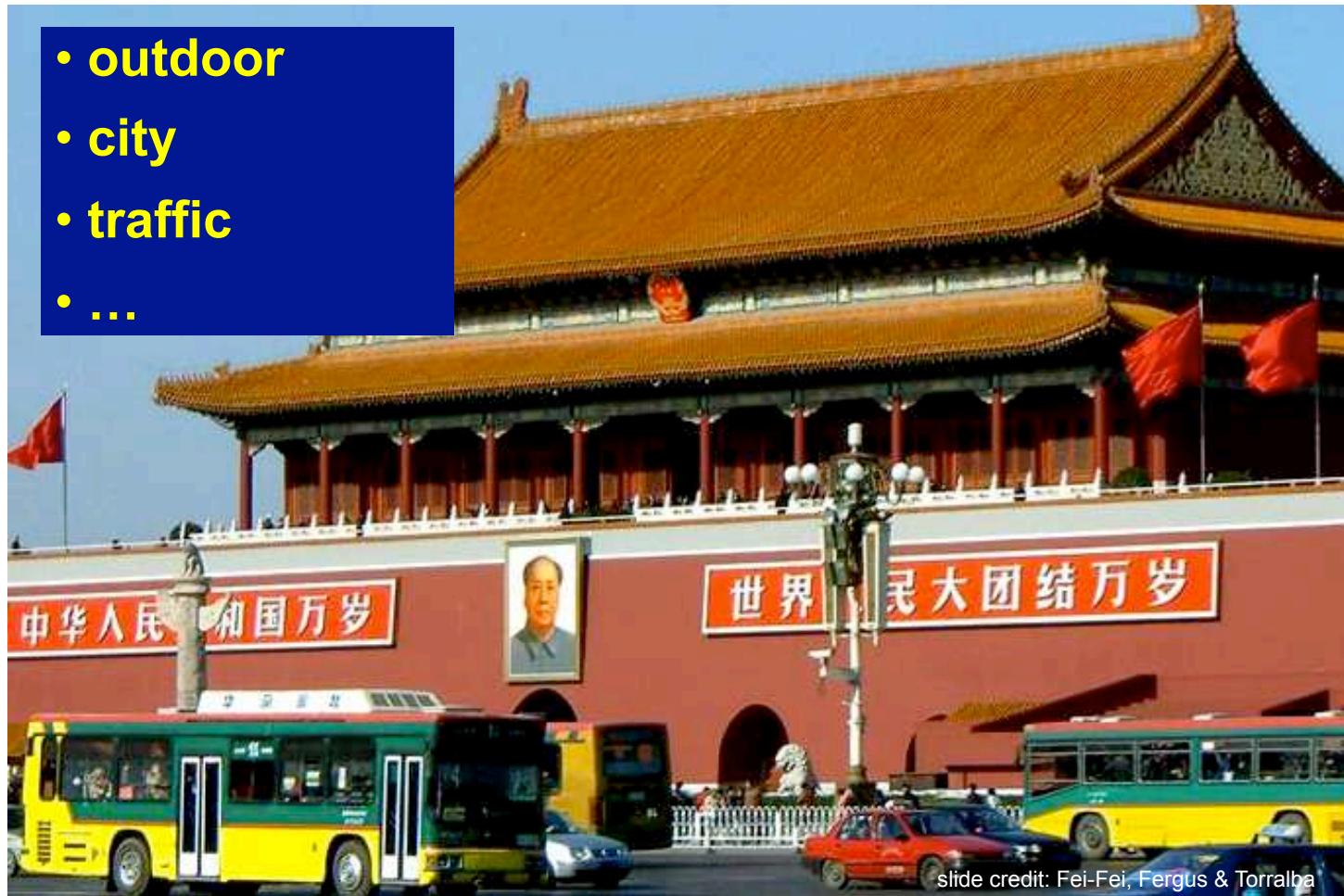
Object categorization



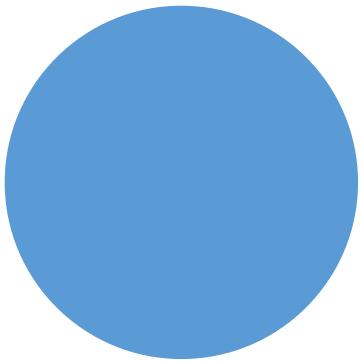
Computer Vision: The very basics

Scene and context categorization

- **outdoor**
- **city**
- **traffic**
- ...



slide credit: Fei-Fei, Fergus & Torralba



Why is Computer
Vision hard? |

Why is Computer Vision hard?

Challenges: Robustness



Illumination



Object pose



Clutter



Occlusions



Intra-class variation



Viewpoint

Why is Computer Vision hard?

Challenges: Robustness



- **Detection in Crowded Scenes**
 - Learn object variability
 - Changes in appearance, scale, and articulation
 - Compensate for clutter, overlap, and occlusion

Why is Computer Vision hard?

Challenges: object intra-class variation



slide credit: Fei-Fei, Fergus & Torralba

Why is Computer Vision hard?

Inherent ambiguity of the problem

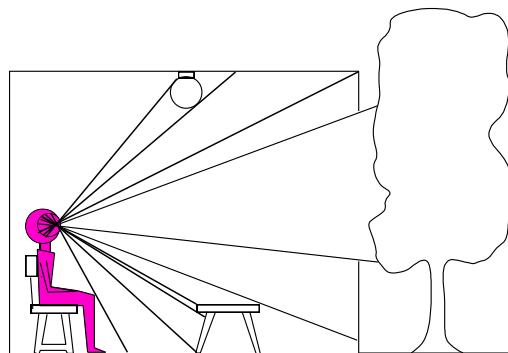
- Many different 3D scenes could have given rise to a particular 2D image



Why is Computer Vision hard?

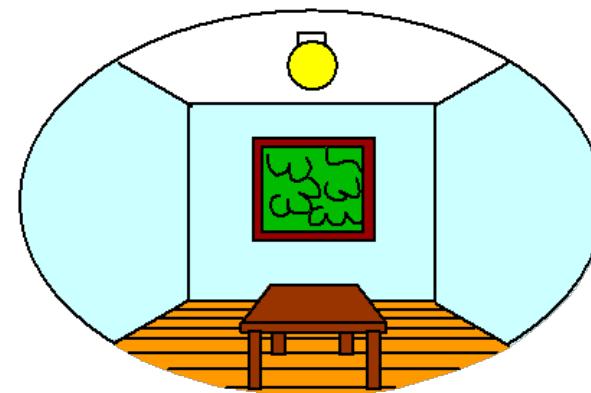
Dimensionality reduction: from 3D to 2D

3D world



Point of observation

2D image

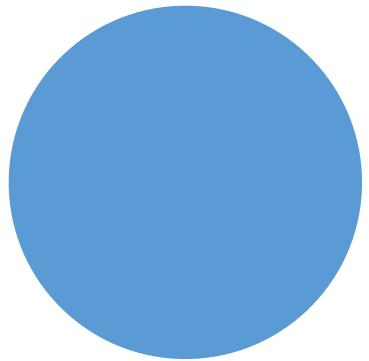


What is preserved?

- Straight lines, incidence

What have we lost?

- Angles, lengths



Computer Vision in the
Personal Robotics Lab

All code available!

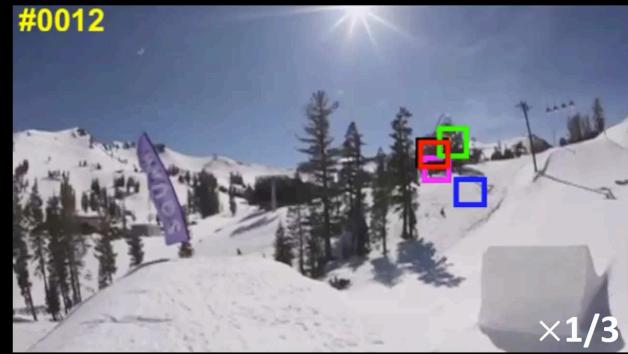
Computer Vision in the Personal Robotics Lab

- Ultra-fast visual object tracking

Attentional Correlation Filter Network
for Adaptive Visual Tracking

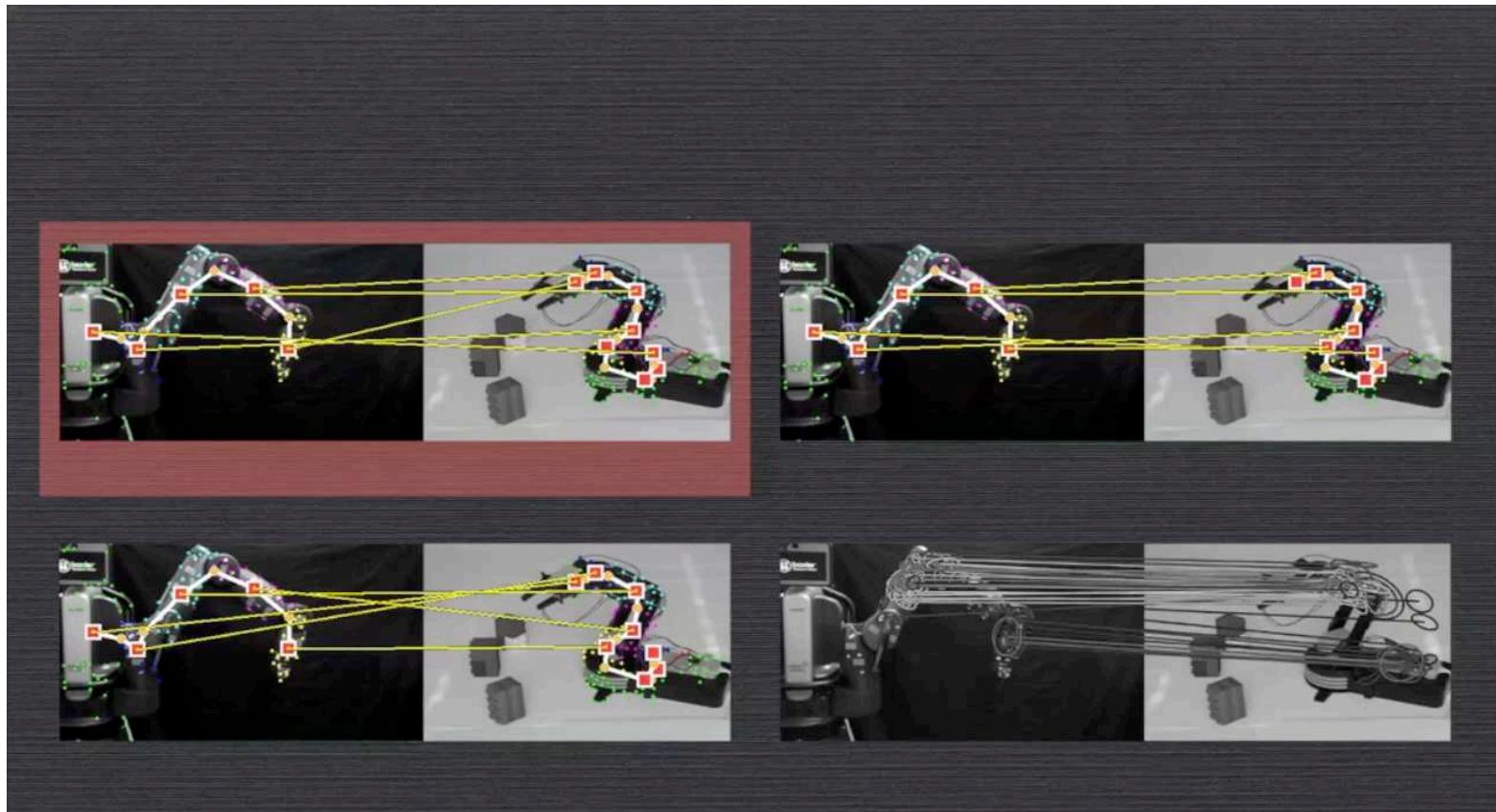
Anonymous #1996

Tracking Result – Skiing



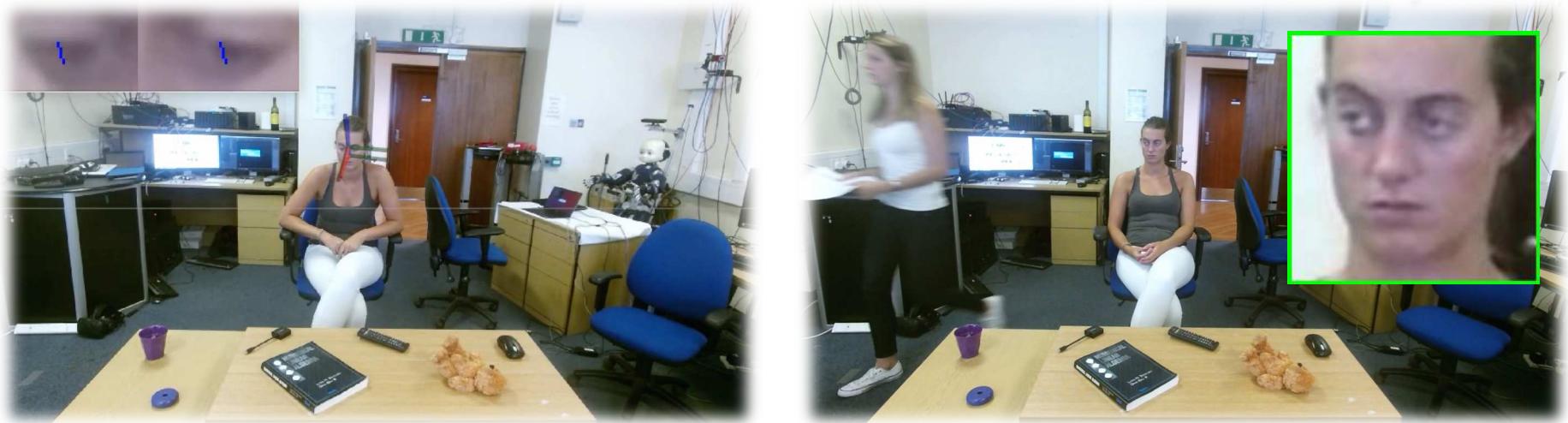
Computer Vision in the Personal Robotics Lab

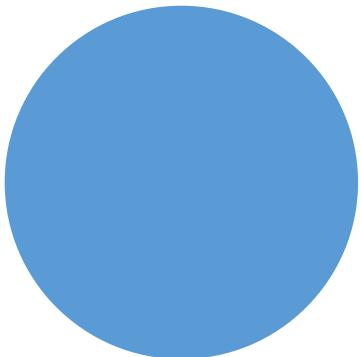
- Robot motion retargeting



Computer Vision in the Personal Robotics Lab

- Gaze & blink estimation





Other popular computer
vision resources

All code available!
Keep in mind you
need powerful GPU

Computer Vision with Tensorflow & PyTorch

- Human pose estimation using Detectron2

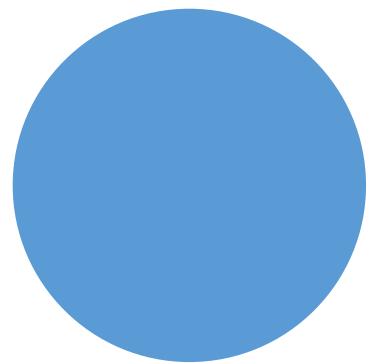


Video credit: Detectron2

Computer Vision with Tensorflow & PyTorch

- Object classification & segmentation using Detectron2

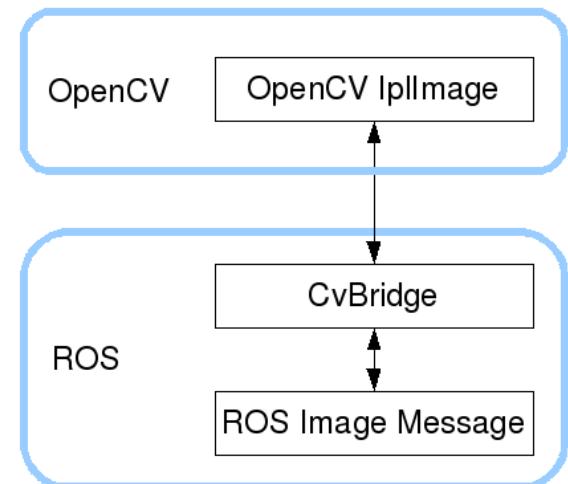




Computer Vision
& ROS |

Computer Vision & ROS

- My recommendation: Python for ease of use
(but we provide some C++ hints as well)
- C++ & Python nodes can interact!
- Most common underlying data structure: *numpy*
- Easy conversion between
ROS *sensor_msgs/Image* &
OpenCV using *cv_bridge*
(for Python & C++)



Acquiring images

- There are different packages to acquire images from a camera
- Webcam: *uvc_camera* or *cv_camera*
- Kinect / Asus Xtion: *openni_launch* <- Provides depth
- If one does not work, try another one ..
wiki.ros.org/Sensors/Cameras

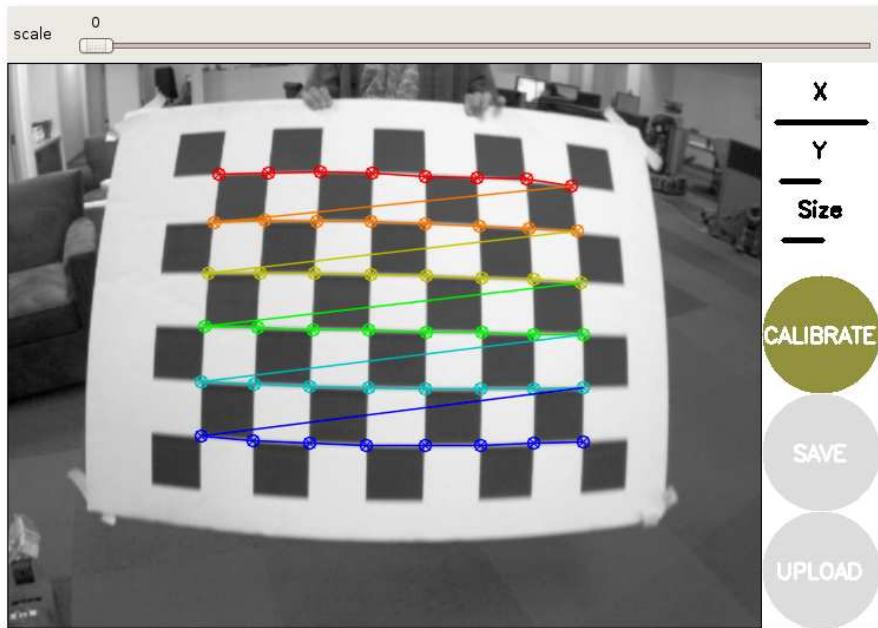
Acquiring images

Unifying data format: sensor_msgs/Image

```
# This message contains an uncompressed image
# (0, 0) is at top-left corner of image
#
Header header          # Header timestamp should be acquisition time of image
                        # Header frame_id should be optical frame of camera
                        # origin of frame should be optical center of camera
                        # +x should point to the right in the image
                        # +y should point down in the image
                        # +z should point into to plane of the image
                        # If the frame_id here and the frame_id of the CameraInfo
                        # message associated with the image conflict
                        # the behavior is undefined
uint32 height           # image height, that is, number of rows
uint32 width             # image width, that is, number of columns
#
# The legal values for encoding are in file src/image_encodings.cpp
# If you want to standardize a new string format, join
# ros-users@lists.sourceforge.net and send an email proposing a new encoding.
string encoding          # Encoding of pixels -- channel meaning, ordering, size
                        # taken from the list of strings in include/sensor_msgs/image_encodings.h
uint8 is_bigendian        # is this data big endian?
uint32 step               # Full row length in bytes
uint8[] data              # actual matrix data, size is (step * rows)
```

Acquiring images

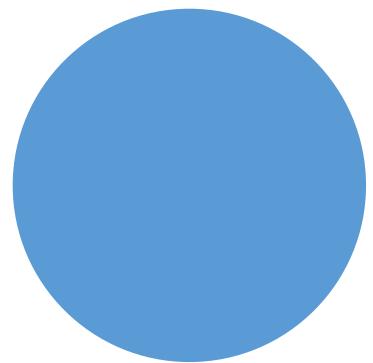
It is often beneficial to make sure your camera is calibrated:
http://wiki.ros.org/camera_calibration/Tutorials/MonocularCalibration



We have a checkerboard pattern in the lab

Pointers to other materials

- <https://github.com/osrf/rosbook>
“Programming Robots with ROS” Chapter 12: Follow-Bot
- [https://industrial-training-master.readthedocs.io/
en/melodic/_source/session5/OpenCV-in-Python.html](https://industrial-training-master.readthedocs.io/en/melodic/_source/session5/OpenCV-in-Python.html)
- ROS Wiki
- OpenCV
- PCL
- Paperswithcode.com



Practical session

Practical session

<https://github.com/Tobias-Fischer/HCR-Vision-Tutorial>

1. Subscribing & publishing images, interfacing with OpenCV
2. Face detection with OpenCV
3. Obtaining 3D coordinates using Kinect-like sensor
Make sure you calibrate colour AND depth cameras:
http://wiki.ros.org/openni_launch/Tutorials/IntrinsicCalibration
http://wiki.ros.org/openni_launch/Tutorials/ExtrinsicCalibration
4. TensorFlow specifics in ROS
5. PyTorch example
6. Brief C++ intro