Let’s talk about Computer Vision!

Silvia Santano and Pepper

Köln, 19.02.2018
About Me

› Silvia Santano
› Pepper Applications development
› At inovex since June 2016
› Programming robots since I was 12
Agenda

› Computer Vision
› Image Recognition
› Pepper
  › Characteristics
  › Computer Vision with Pepper
  › External services
    – Google
    – Microsoft
› On-device CV with CNNs
Agenda

› Computer Vision
  › Image Recognition
  › Pepper
    › Characteristics
    › Computer Vision with Pepper
  › External services
    - Google
    - Microsoft
  › On-device CV with CNNs
Computer Vision

Automatic extraction, analysis and understanding of information from images
Humans can recognize objects in images with little effort despite of huge variations.

For computers this is still a challenge...
Agenda

› Computer Vision
› Image Recognition
› Pepper
   › Characteristics
   › Computer Vision with Pepper
   › External services
     – Google
     – Microsoft
› On-device CV with CNNs
Image Recognition

Determine whether or not an image contains a specific object
Image Recognition

Main subfields:

› Classification
› Object detection
› Semantic segmentation
› Identification
Image Recognition

Main subfields:

› Classification
› Object detection
› Semantic segmentation
› Identification
Image Recognition

Main subfields:

› Classification
› Object detection
› Semantic segmentation
› Identification

Images: http://david.grangier.info/scene_parsing/
Image Recognition

Main subfields:

› Classification
› Object detection
› Semantic segmentation
› Identification
Agenda

› Computer Vision
› Image Recognition
› Pepper
   › Characteristics
   › Computer Vision with Pepper
   › External services
     - Google
     - Microsoft
› On-device CV with CNNs
Pepper: Capabilities

- Softbank Robotics
- Voice and gestures
- Face and emotion recognition
- Internet Connection
- Tablet
- Safety mechanisms, automatic balance, and anti-collision system
**Pepper: Technical Characteristics (v1.8A)**

<table>
<thead>
<tr>
<th>PROCESSOR</th>
<th>Atom E3845</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>Quad core</td>
</tr>
<tr>
<td>Clock speed</td>
<td>1.91 GHz</td>
</tr>
<tr>
<td>RAM</td>
<td>4 GB DDR3</td>
</tr>
<tr>
<td>OS</td>
<td>Nao QI OS</td>
</tr>
</tbody>
</table>

2 HD Cameras (OV5640)  
1 3D Sensor (ASUS XTION)  
4 Microphones  
A 3-axis Gyrometer and a 3-axis Accelerometer  
6 laser line generators  
2 Infra-Red sensors  
2 ultrasonic sensors  
3 tactile sensors  
3 bumpers  
20 Motors and actuators

<table>
<thead>
<tr>
<th>Tablet</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions</td>
</tr>
</tbody>
</table>
| CPU | 1.3 GHz quad-core ARM Cortex-A7  
Cache 512 KB L2  
Wi-Fi, Bluetooth  
1.6G pixel/sec @416MHz |
| DDR3 SDRAM | 1GB (512MB * 2) |
| Flash Memory | 32GB (eMMC) |
| Display | Type: IPS, Resolution: 1280*800  
Color: 24bit true color |
| Touch Panel | Capacitive Multi-Touch (5 point) |
| Sensors | Light illumination, Acceleration  
Gyro, Geomagnetic |
| OS | Android (6.0) |
Programming Pepper

› Choregraphe und Python
› Python
› C++
› Javascript
› Soon: Android (reduced function set)
Programming Pepper

Choregraphe
# Volume

```
#can_you
"ein bisschen" etwas leiser [sprechen reden]
"sprich {"ein bisschen" etwas} leiser"
"Dreh die Lautstärke runter"
"sprich nicht so laut"
"du sprichst zu laut"
```

\^gotoReactivate\(decrease\_volume\)
\u:($empty) %decrease\_volume
\^call(ALVolumeSlider.decreaseVolume()) $Demo/back=1
\c1:(false) es tut mir leid, das ist das Minimum
\c1:(true) okay ich spreche jetzt leiser

```
\u2:([
  nochmal
  mehr
  "noch {"ein bisschen" etwas} mehr"
  "immer noch zu laut"
])
\^gotoReactivate\(decrease\_volume\)
Agenda

› Computer Vision
› Image Recognition
› Pepper
    › Characteristics
    › Computer Vision with Pepper
› External services
    – Google
    – Microsoft
› On-device CV with CNNs
Computer Vision with Pepper

› Face Detection and People Tracking
› Face Learning and Recognition
› People Characteristics Perception
Computer Vision with Pepper
People Characteristics Perception

PeoplePerception/Person/<ID>/AgeProperties
PeoplePerception/Person/<ID>/ExpressionProperties
PeoplePerception/Person/<ID>/GenderProperties
PeoplePerception/Person/<ID>/SmileProperties
PeoplePerception/Person/<ID>/FacialPartsProperties
PeoplePerception/Person/<ID>/Distance
PeoplePerception/Person/<ID>/IsFaceDetected
PeoplePerception/Person/<ID>/IsVisible
PeoplePerception/Person/<ID>/NotSeenSince
PeoplePerception/Person/<ID>/PresentSince
PeoplePerception/Person/<ID>/RealHeight
PeoplePerception/Person/<ID>/ShirtColor
Computer Vision with Pepper

› Face Detection and People Tracking
› Face Learning and Recognition
› People Characteristics Perception
› Emotion Recognition
Computer Vision with Pepper

Emotion Recognition Module

Data sources:
- Expression and smile
- Acoustic voice emotion analysis
- Head angles
- Touch sensors
- Semantic analysis from speech
- Sound level and energy level of noise
- Movement detection

Valence
Attention Level
Smile
Expression
{
  "calm",
  "anger",
  "joy",
  "sorrow",
  "laughter",
  "excitement",
  "surprise"
}
(Real values normalized)
Computer Vision with Pepper

› People Tracking
› Face Detection, Learning and Recognition
› People Perception
› Emotion Recognition
› Vision Recognition
› Barcode Reader
Computer Vision with Pepper

DEMO:

People Perception
Emotion Recognition
Face Detection and Recognition
Agenda

› Computer Vision
› Image Recognition
› Pepper
  › Characteristics
  › Computer Vision with Pepper
  › External services
    - Google
    - Microsoft
› On-device CV with CNNs
External services integration

Google’s Machine Learning Cloud Vision API

› Machine learning service with pre-trained models
› JSON REST API + client libraries (C#, GO, Java, Node.js, PHP, Python, Ruby)

Explicit Content Detection
Logo Detection
Label Detection
Landmark Detection
Optical Character Recognition
Face Detection
Image Attributes
Web Detection
External services integration
Google’s Machine Learning Cloud Vision API: LABELS

https://cloud.google.com/vision/
External services integration

Google’s Machine Learning Cloud Vision API: LOGO DETECTION
External services integration
Google’s Machine Learning Cloud Vision API: LABELS

Show me the code
External services integration
Google’s Machine Learning Cloud Vision API: LABELS

```python
def detect_labels(path):
    """Detects labels in the file."""
    client = vision.ImageAnnotatorClient()

    with io.open(path, 'rb') as image_file:
        content = image_file.read()

    image = types.Image(content=content)

    response = client.label_detection(image=image)
    labels = response.label_annotations
    print('Labels:
    for label in labels:
        print(label.description)
```

client libraries (C#, GO, Java, Node.js, PHP, Python, Ruby)
External services integration
Google’s Machine Learning Cloud Vision API: LABELS

POST https://vision.googleapis.com/v1/images:annotate?key=YOUR_API_KEY

```json
{
  "requests": [
    {
      "image": {
        "content": "<base64-encoded-image-content>"
      },
      "features": [
        {
          "type": "LABEL_DETECTION"
        }
      ]
    }
  ]
}
```

JSON REST API
External services integration
Google’s Machine Learning Cloud Vision API

DEMO:
Logo Detection
Label Detection
Optical Character Recognition
Web Detection
Emotion Detection
External services integration
Microsoft Cognitive Services

› Machine learning service with pre-trained models
› JSON REST APIs + client libraries (C#, Android, Swift)

Computer Vision API:
  Analyze Image
  Optical Character Recognition
  Handwritten Text Detection
Face API
Emotion API
External services integration
Microsoft Cognitive Services: COMPUTER VISION API

https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/
External services integration
Microsoft Cognitive Services: FACE API

https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/
External services integration

Microsoft Cognitive Services

DEMO:

Analyze Image
Optical Character Recognition
Handwritten Text Recognition
Emotion Detection
Agenda

› Computer Vision
› Image Recognition
› Pepper
   › Characteristics
   › Computer Vision with Pepper
   › External services
     – Google
     – Microsoft
› On-device CV with CNNs
On-device CV with CNNs

Why

› Privacy
› Latency
› Connectivity
› Security
› Cost
On-device CV with CNNs

Limitations

› Compute
› Memory
› Storage
› Power
› Bandwidth
On-device CV with CNNs

Tools

- e.g. Tensorflow Mobile
- or Tensorflow Lite

Pre-trained models
Tensorflow Object Detection API
...Out of curiosity

Google’s Algorithm found houses

Google’s Algorithm found no houses
Yes, but: chihuahua or muffin?
Comparison

› Amazon’s Rekognition is not just good at identifying the primary object but also the many objects around it
› Google’s Vision API and IBM Watson Vision return straightforward, descriptive labels
› Microsoft’s tags were usually too high level
› Clarifai returns, by far, the most tags (at 20) although very generic tags. It also adds qualitative and subjective labels, such as “cute”, “funny”, “adorable”, and “delicious”
Vielen Dank

Silvia Santano
Application Development

@SilviaSantano
linkedin.com/in/silviasantano
ssantano@inovex.de
0173 3181 085