Computer Vision for HCI –
From Smart Rooms to Vision for the Blind

Rainer Stiefelhagen
Institute for Anthropomatics and Robotics, KIT &
Study Center for Visually Impaired Students, KIT
Multimodal Interfaces in Alex` Labs!
(an incomplete snapshot)

- Alex also a pioneer in multimodal interfaces
  - Face tracking
    - Hunke & Waibel 94, Yang & Waibel 96/98)
    - 3 papers > 1500 citations
  - Lip-reading (from 1992/93)
    - Bregler, Hild, Manke, Duchnowski, Meier, Yang,
  - Emotions in speech
    - Dellaert, Polzin, Waibel 96 (>550 citations)
  - Handwriting, NPEN, NPEN++ (from 1994)
    - Manke, Finke, Jäger, Reichert, Bodenhausen …
      (>200 citations)
  - Head pose and attention
    - Schiele & Waibel 1995
  - Multimodal dialogue, error repair, …
Multimodal Interfaces in Alex` Lab(s) (2)

- Computer-Vision for HCI team in KA (2001-2009)
  - Tracking (Bernardin), Gestures (Nickel), Face ID (Ekenel, Fischer, Gao), Head pose & attention (Voit)
- Multimodal human-robot interaction
  - SFB 588 Humanoid Robots (2001-2012)
  - AV-Tracking, face-ID, pointing gestures, gaze, speech + dialogue
- Context-Aware Rooms & Services
  - The Connector – Context Aware Smart Office
  - Smart seminar & meeting rooms
  - CLEAR Evaluations together with NIST
Now: Computer Vision for HCI Lab

- Computer Vision for HCI Lab at KIT
  (+ Perceptual User Interfaces Lab at Fraunhofer IOSB)
  - Visual perception of humans: tracking, identification, age, gender, facial expression, action recognition, …
  - Interactive systems: Smart control rooms, driver monitoring, smart OP, sleeping labs / nursing homes, interaction with people with dementia
  - Movie-analysis, video-surveillance

- Study Center for Visually Impaired Students (SZS)
  - Supports blind and partially sighted students at KIT
  - Provides accessibility to all courses, lecture materials, exams, math, software, …

- Mission: overcome barriers for the seeing impaired
Problem

- Ca. 285 Mio. visually impaired worldwide (~39 Mio. blind)
- Loss of sight leads to many difficulties
  - Orientation & mobility
  - Communication & social interaction
  - Information access, university, job
  - Activities of daily living (cooking, clothing, money, medicine, applicances, …)

- E.g. mobility:
  - Three out of ten visually impaired never leave their homes alone (for elderly: 2 out of 3), 30% try to avoid unknown routes
  - Reason: fear of accidents and orientation problems

- Interaction
  - Recognize who is there, who is looking at me, etc.

- **Computer Vision gives us new ways to assist!**
Recognizing persons

- Person detection and tracking
  - Better, faster systems, already in use in cars

- Person identification,
- Age, gender, facial expression recognition

Can be used
- to support social interaction
- to avoid collisions
Object Recognition / ImageNet Challenge

- Very fast progress in object recognition
- E.g. Imagenet
- Top 5 error for 1000 class now < 5%
- → on par with human

Methods:
- CNNs (of course)
Image & Video Captioning

A person riding a motorcycle on a dirt road.
Two dogs play in the grass.
A yellow school bus parked in a parking lot.

"black and white dog jumps over bar."
"young girl in pink shirt is swinging on swing."

Methods: CNN + RNN
CNNs + RNNs
**Image & Video Question-Answering**

**LOTR: Return of the King**

**Why does Arwen wish to stay in Middle Earth?**

- She is too weak to travel
- She wants to die on Middle Earth
- Her son asked her to stay
- **Arwen sees her son in her visions**
- She likes Middle Earth

---


What are we going to do with it?

- **Mobility and navigation support**
  - Obstacles, crossings, buildings, entrances

- **Social interaction support**
  - Who is there, who is looking at me?

- **Accessibility of lecture material**
  - images, diagrams, charts, …

- **How to design the interface?**
  - Audio-tactile interfaces
**TERRAIN: Mobility and Navigation support**

- Funded by BMBF: July 2016 to June 2019
- Goal: Build a vision based assistive system for urban mobility

- Detect and tell (Computer Vision)
  - Free way vs. obstacles
  - Street lights and crossings
  - Specific buildings and entrances

- Have to investigate user interface (HCI)
  - How to tell directions, obstacles, scene information, etc.
  - Audio-tactile interfaces
  - Braille-interface
  - Smartphone as front-end
First steps: Accessible section detection

Accessible Section Detection for Visual Guidance

MAP4VIP@ICME2013
3FPS with labeled ground truth

True Positive  True Negative
False Positive False Negative

Accessible Section Detection for Visual Guidance

MAP4VIP@ICME2013
3FPS with labeled ground truth

True Positive  True Negative
False Positive False Negative
Zebra-Crossing Detection

- Important for safe street crossing
- In aerial images
  - To enhance Open Street Map data

- Also from mobile cam

<table>
<thead>
<tr>
<th>Method</th>
<th>Precision</th>
<th>Recall</th>
<th>Accuracy</th>
<th>Avg.-Prec.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOG^{30x30}-lin</td>
<td>74.8</td>
<td>93.1</td>
<td>92.4</td>
<td>94.43</td>
</tr>
<tr>
<td>HOG^{20x20}-RBF</td>
<td>95.2</td>
<td>96.2</td>
<td><strong>98.9</strong></td>
<td>97.99</td>
</tr>
<tr>
<td>LBP^{17/10}-lin</td>
<td>99.4</td>
<td><strong>97.4</strong></td>
<td>98.4</td>
<td><strong>99.56</strong></td>
</tr>
<tr>
<td>LBP^{17/10}-RBF</td>
<td><strong>99.7</strong></td>
<td>97.0</td>
<td>98.3</td>
<td><strong>99.56</strong></td>
</tr>
</tbody>
</table>

Koester & Stiefelhagen, ICCHP 2016
Traffic lights / buttons / crossings

- Collected and annotated training set

- Evaluated various features & classifiers
  - HOG, LBP, LBPH, DeCaf, SVM, boosting, saliency, …

- Crop-classification works well
  - Walklight: 0.97 (AUC)
  - Buttons: 0.95 (AUC)

- However, not yet good enough for real application (detection)
  - Too many FP
  - Further work needed …
Tactile vs. Audio output for obstacle avoidance

- Obstacle course 20 x 5m
- Audio vs. Tactile feedback on white cane
- Six users: 3 blind, 3 vis. impaired
- Evaluation using NASA TLX

- Blind users (trained white cane users) strongly preferred haptic output
- Partially sighted users preferred audio
Positioning of vibro-tactile actuators

- Wizard-of-Oz experiments
  - Find a book in the cabinet
  - Navigate through some obstacles

- Placement on palm below finger works well, also back-side between thumb and pointing finger

- Vibratation-feedback along direction to go seems more intuitive than indicating the obstacle

- Tactile belt
Social interaction support

- BMBF funded project AVVIS
- Tell blind user who is there, who is looking at him / her
- Computer vision modules ready
- Now investigating audio-tactile output variants

- Motivated by a survey: 24 out of 25 users
  - would like to get visual information about other person
  - Would like to know when other person is looking at him/her
  - Would like to get info about facial expressions
Some more things

- Tram door detection
- Vehicle sonification
- Wall plug detection
- Button detection & reco.
- Playing cards ...
- Dice ...

CV:HCl 2013 Praktikum Results

1. Color segmentation
2. Edges and corners
Summary

- Computer vision gives new possibilities to build AT for the visually impaired
  - Rapid progress in computer vision

- Need to design and adapt the interface
  - Audio: speech, sounds, earcons, …
  - Vibro-Tactile: positioning, which patterns, …
  - Provide flexibility

- Important to work with your blind / vip user groups
  - (not blinfolded Ph.D. students!)

- Things need to be integrated, robust, fast
- Dream: have everything run on your smart phone (affordable & cool)
Acknowledgements

CVHCI/SZS
- D. Koester, M. Martinez, Monica Haurilet, S. Sarfraz, Boris Schauerte, M. Tapaswi, T. Wörwein (CVHCI)
- Angela Constantinescu, Gerhard Jaworek, Karin Müller, Thorsten Schwarz (SZS)

Smart rooms & cars, PUI @ IOSB
- M. Voit, F. van de Camp, M. Martin

Funding by
- Google: Research Faculty Award
- DFG, BMBF: Projects AVVIS, TERRAIN, KonsensOP,
- KIT, Fraunhofer

Thanks a lot and congratulations to Alex!