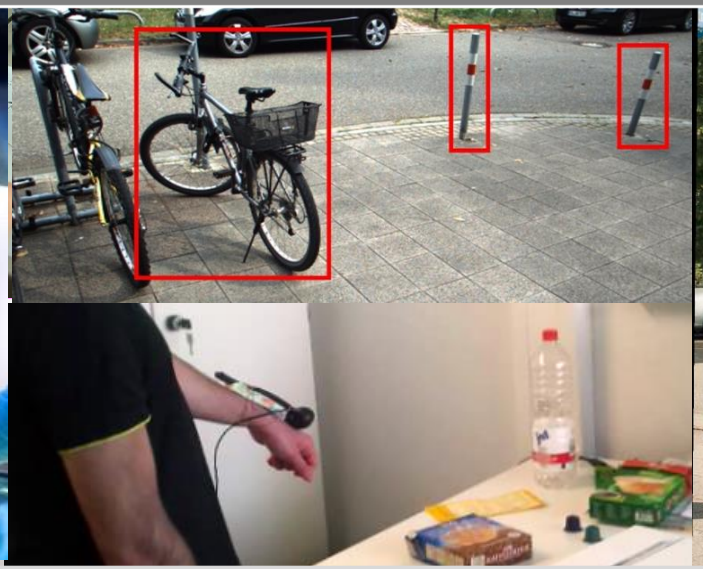
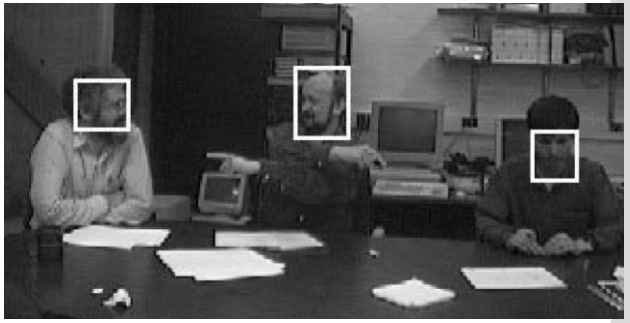
Computer Vision for HCI –
From Smart Rooms to Vision for the Blind

Rainer Stiefelhagen
Institute for Anthropomatics and Robotics, KIT &
Study Center for Visually Impaired Students, KIT

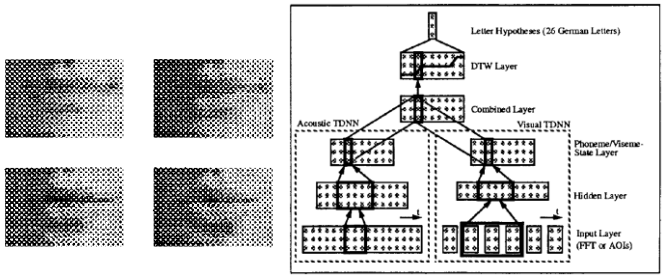


Multimodal Interfaces in Alex` Labs !
(an incomplete snapshot)

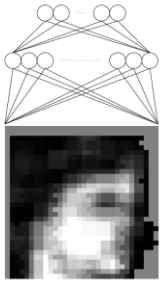
- Alex also a pioneer in multimodal interfaces
- Face tracking
  - Hunke & Waibel 94, Yang & Waibel 96/98)
  - 3 papers > 1500 citations
- Lip-reading (from 1992/93)
  - Bregler, Hild, Manke, Duchnowski, Meier, Yang,
  - My first Hiwi-Job 1993 & BA-thesis 1995
- Emotions in speech
  - Dellaert, Polzin, Waibel 96 (>550 citations)
- Handwriting, NPEN, NPEN++ (from 1994)
  - Manke, Finke, Jäger, Reichert, Bodenhausen … (>200 citations)
- Head pose and attention
  - Schiele & Waibel 1995
  - My master thesis (´96), doctoral thesis (2002)
- Multimodal dialogue, error repair, …



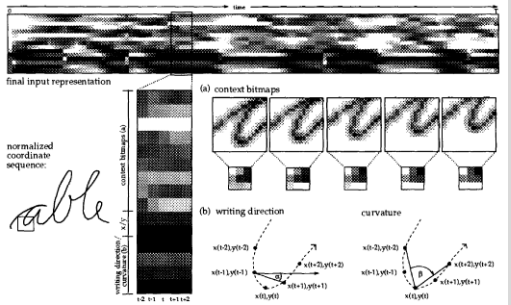
Faces: Yang & Waibel 96



AV-ASR: Bregler, Manke, Hild, Waibel 93



Schiele & Waibel 95



Manke, Finke, Waibel, NIPS 1995

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
  - EU FP7: Computers in the Human Interaction Loop – CHIL (2004-2007)
  - The Connector – Context Aware Smart Office
  - Smart seminar & meeting rooms
  - CLEAR Evaluations together with NIST

Now: Computer Vision for HCI Lab

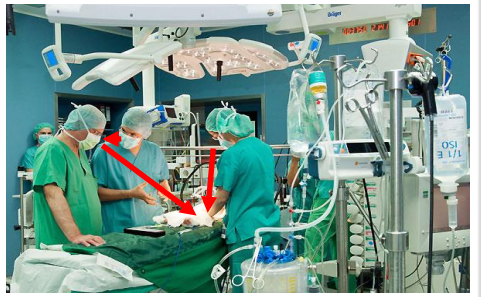
- Computer Vision for HCI Lab at KIT

  (+ 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 partically sighted students at KIT
  - Provides accessibility to all courses, lecture materials, exams, math, software, …

- Mission: overcome barriers for the seeing impaired



Problem

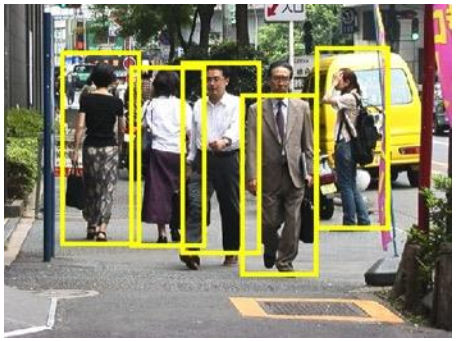
- Ca. 285 Mio. visually impaired worldwide (~39 Mio. blind)
- Loss of sight leads to many difficulties
  - Orientation & mobility
  - Communication & social interaction
  - Information access, university, job
  - Activities of daily living (cooking, clothing, money, medicine, applicances, …)

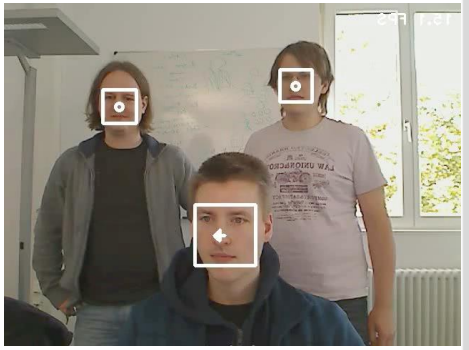
- E.g. mobility:
  - Three out of ten visually impaired never leave their homes alone (for elderly: 2 out of 3), 30% try to avoid unknown routes
  - Reason: fear of accidents and orientation problems
- Interaction
  - Recognize who is there, who is looking at me, etc.

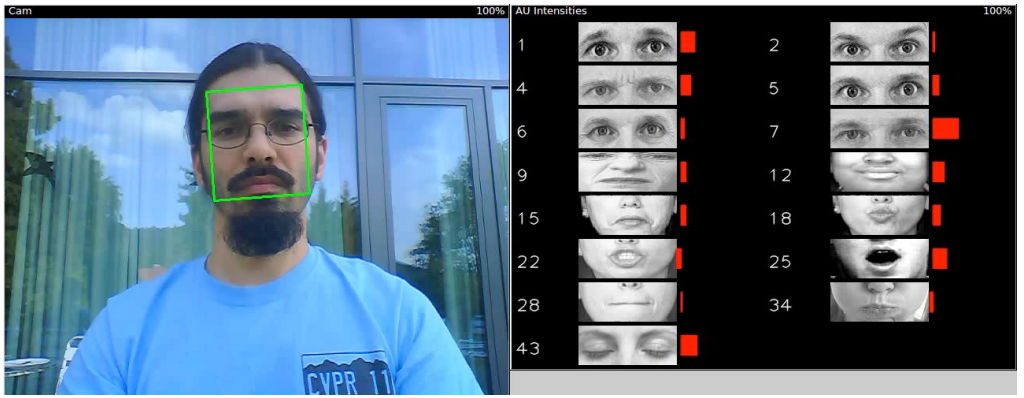
- <u>Computer Vision gives us new ways to assist!</u>

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)

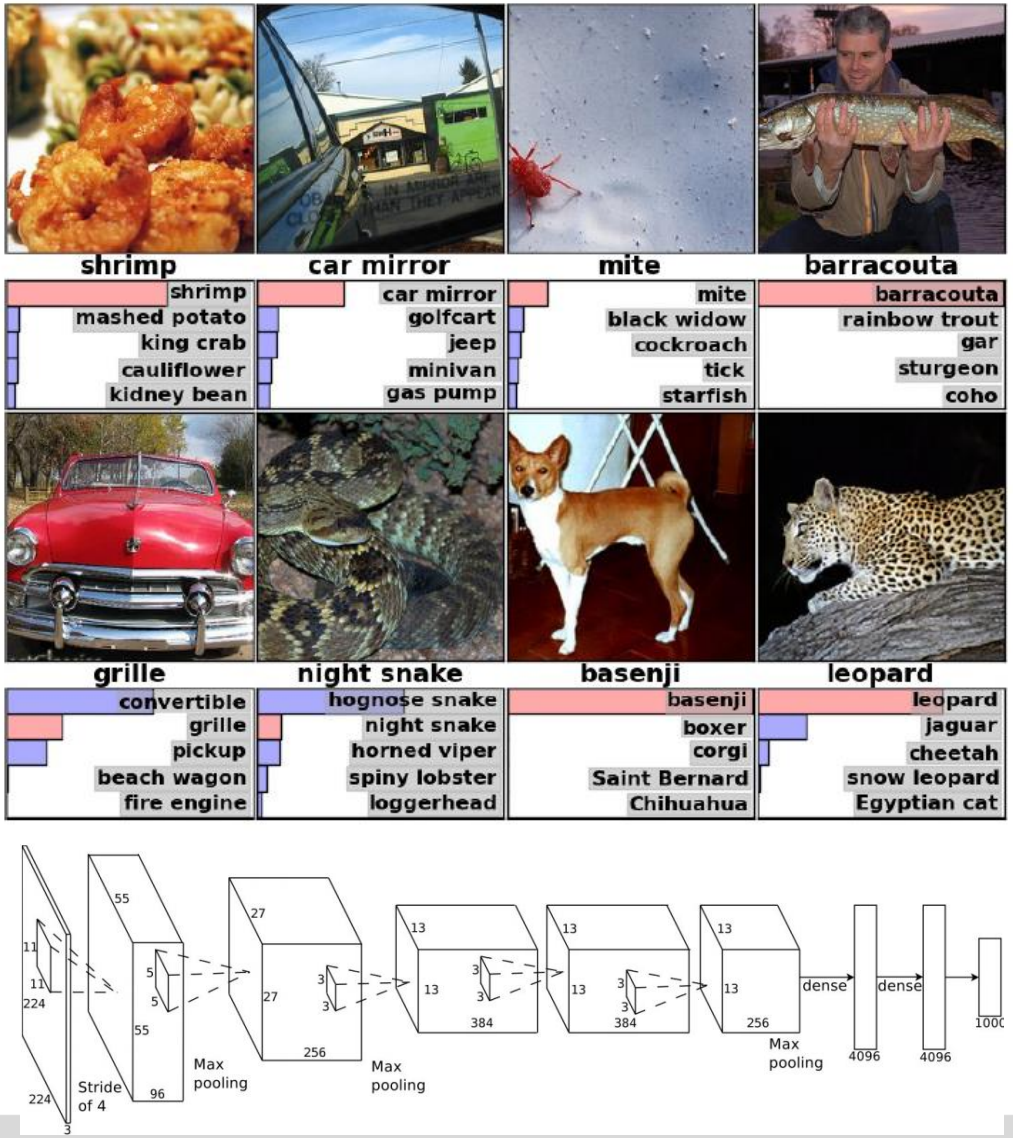


Image & Video Captioning

A person riding a motorcycle on a dirt road.

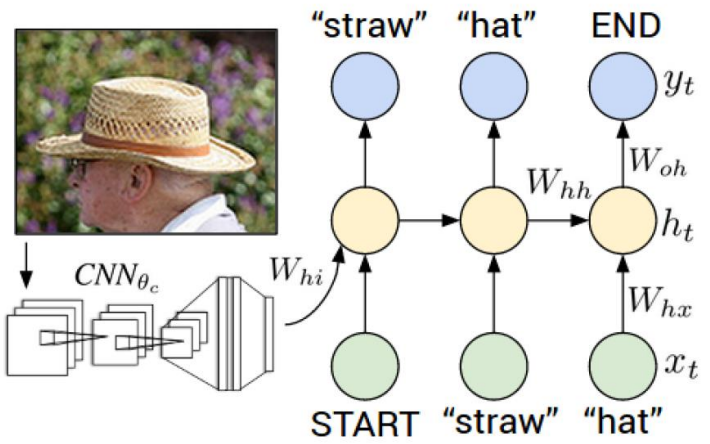
Two dogs play in the grass.

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

What color are her eyes?
What is the mustache made of?

Does it appear to be rainy?
Does this person have 20/20 vision?

S. Antol, et al. VQA: Visual Question Answering. In *ICCV* 2015.

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

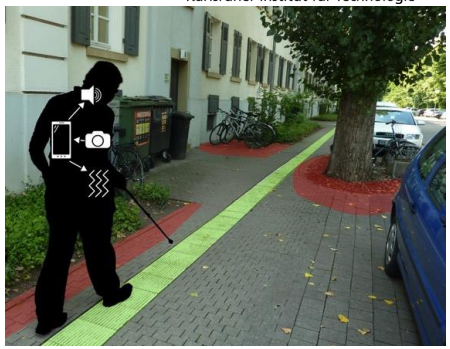
Arwen sees her son in her visions

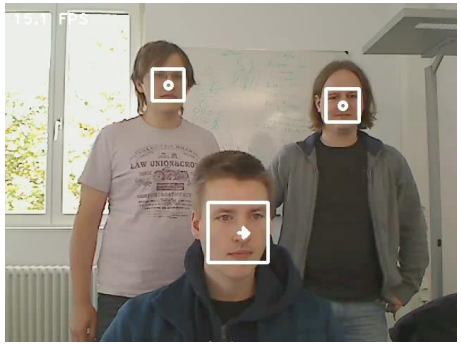
She likes Middle Earth

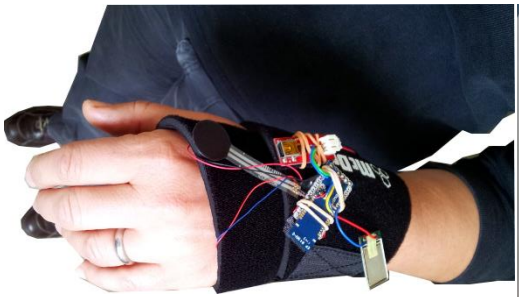
M. Tapaswi et al. MovieQA, In *CVPR* 2016.

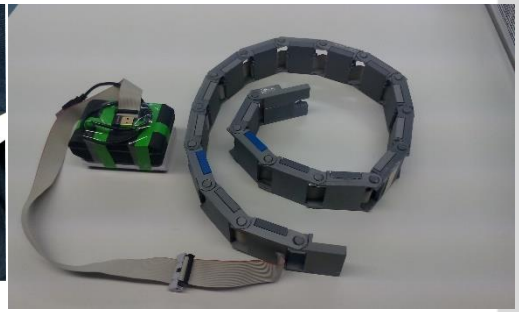
What are we going to do with it?

- 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



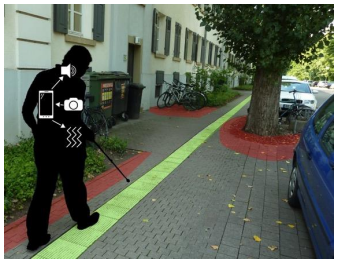




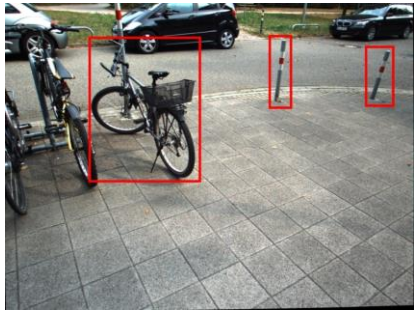


TERRAIN: Mobility and Navigation support

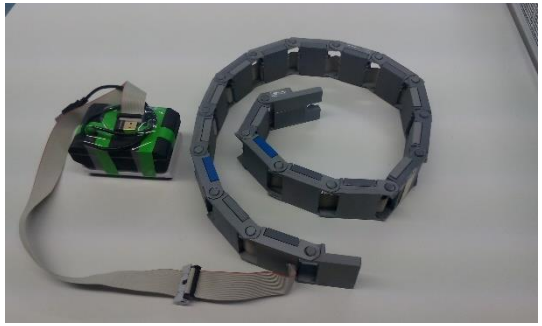
- Funded by BMBF: July 2016 to June 2019
- Goal: Build a vision based assistive system for urban mobility



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

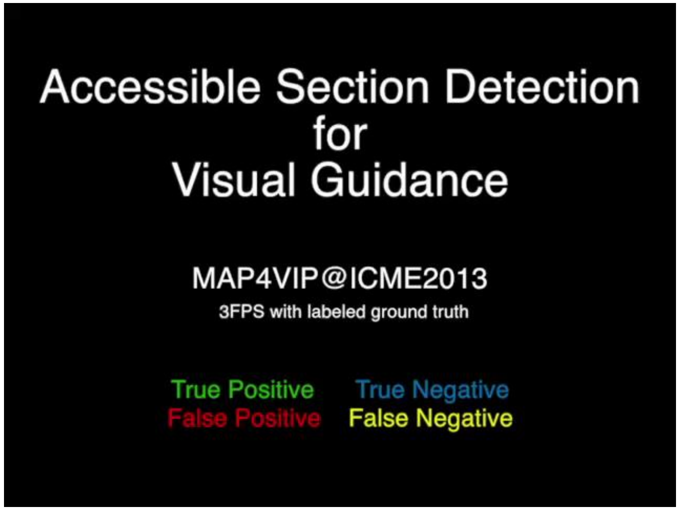


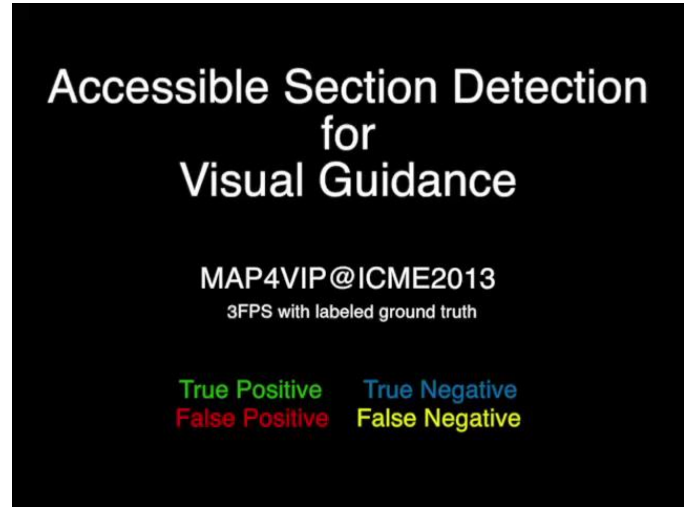
- Have to investigate user interface (HCI)
  - How to tell directions, obstacles, scene information, etc.
  - Audio-tactile interfaces
  - Braille-interface
  - Smartphone as front-end

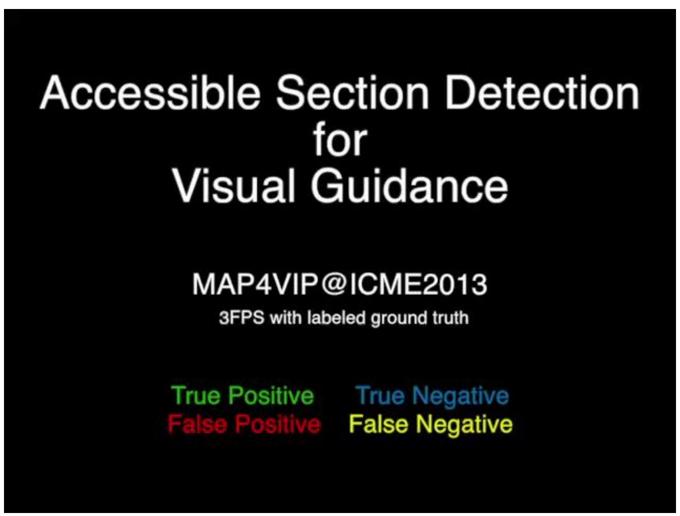


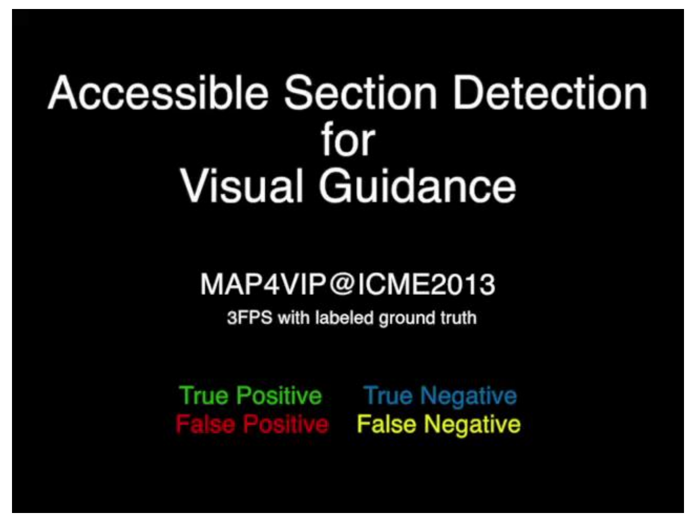
First steps: Accessible section detection

Karlsruher Institut für Technologie





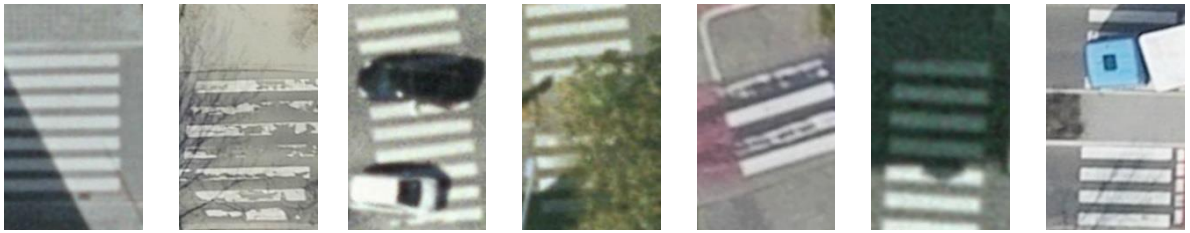




Zebra-Crossing Detection

- Important for safe street crossing

- In aerial images
  - To enhance Open Street Map data



- Also from mobile cam

Method	Precision	Recall	Accuracy	Avg.-Prec.
$HOG^{30\times30}$-lin	74.8	93.1	92.4	94.43
$HOG^{20\times20}$-RBF	95.2	96.2	**98.9**	97.99
$LBP^{17/10}$-lin	99.4	**97.4**	98.4	**99.56**
$LBP^{17/10}$-RBF	**99.7**	97.0	98.3	**99.56**

Koester & Stiefelhagen, ICCHP 2016

Traffic lights / buttons / crossings

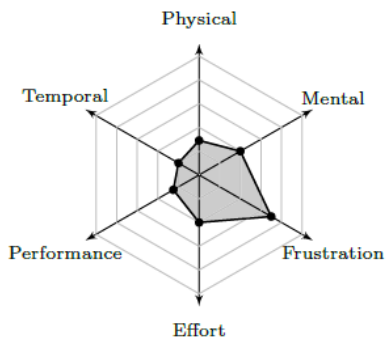
- Collected and annotated training set
- Evaluated various features & classifiers
  - HOG, LBP, LBPH, DeCaf, SVM, boosting, saliency, …
- Crop-classification works well
  - Walklight: 0.97 (AUC)
  - Buttons: 0.95 (AUC)
- However, not yet good enough for real application (detection)
  - Too many FP
  - Further work needed …

walklights buttons

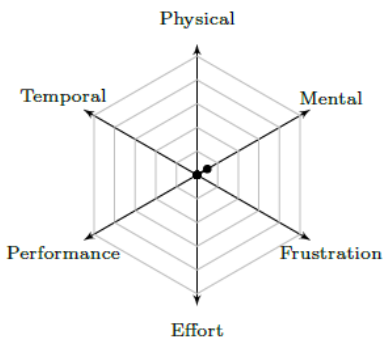
Tactile vs. Audio output for obstacle avoidance

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

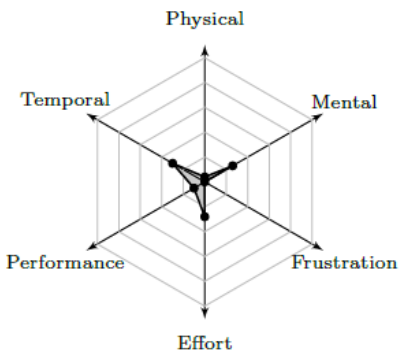
- Blind users (trained white cane users) strongly preferred haptic output
- Partially sighted users preferred audio



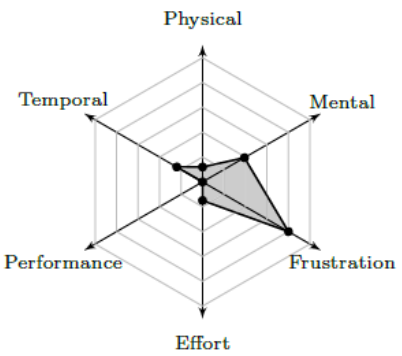
(a) White Cane Users - Audio



(b) White Cane Users - Haptic



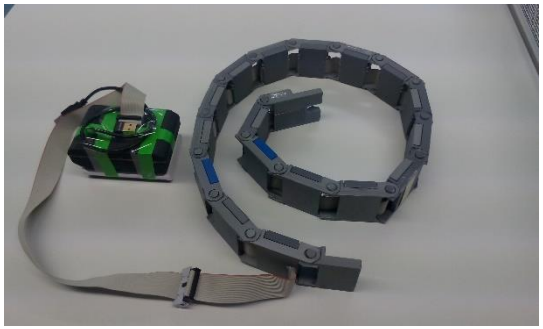
(c) Non White Cane Users - Audio

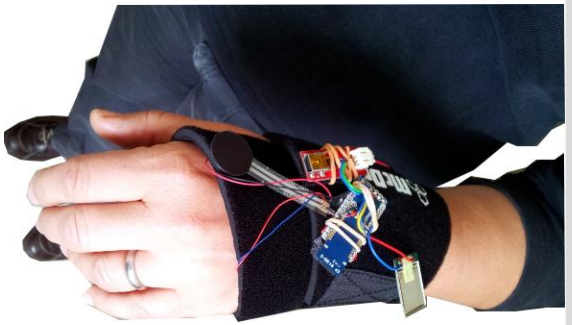


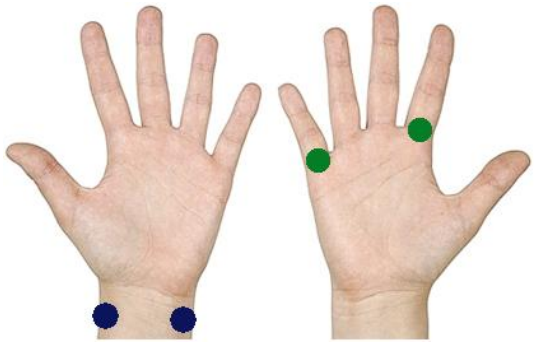
(d) Non White Cane Users - Haptic

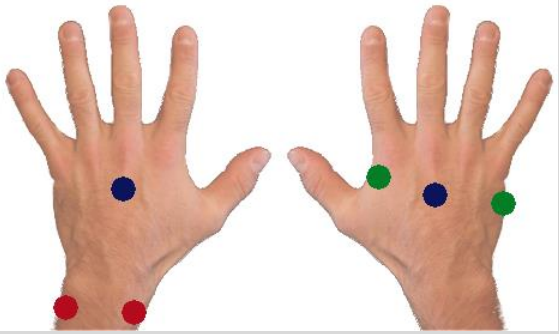
Positioning of vibro-tactile actuators

- 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 mor intuitive than indicating the obstacle
- Tactile belt





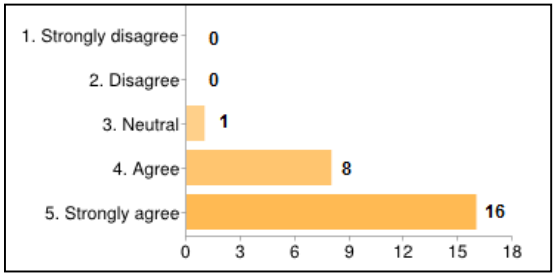


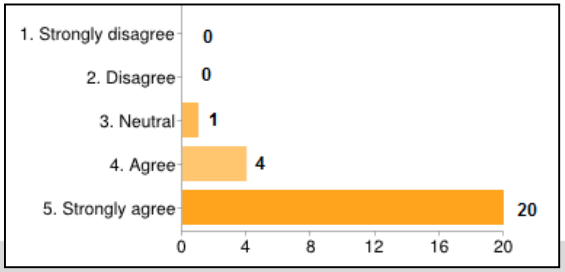


Social interaction support

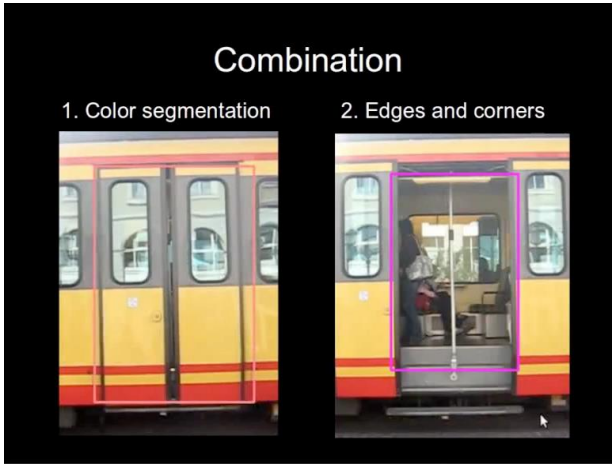
- BMBF funded project AVVIS
- Tell blind user who is there, who is looking at him / her
- Computer vision modules ready
- Now investigating audio-tactile output variants

- 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

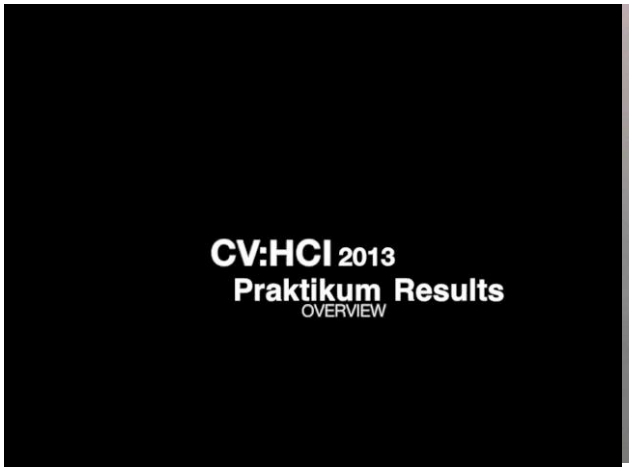


Tram door detection

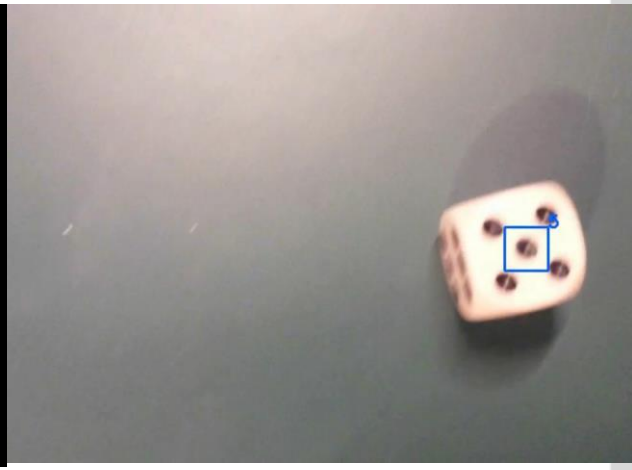
Vehicle sonification

Wall plug detection

Button detection & reco.



Playing cards …



Dice …

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!**