Mobile Information Systems

Lecture 05: I/O on mobile devices (2/2)

© 2015-24 Dr. Florian Echtler
Bauhaus-Universität Weimar
Aalborg University

I/O on mobile devices

- Issues, revisited
  - Last week: Touch, gestures, motion
  - Today: Vision, speech & other I/O channels
- Solutions
  - Common (commercial) approaches
  - Research projects

I/O issues: bimanual (recap)

Image source (CC): https://en.wikipedia.org/wiki/Text_messaging#/media/File:Texting.jpg

- Symmetric
  - Both hands have same role, e.g. typing with both thumbs
- Asymmetric
  - Hands have different roles, e.g. one hand holds device, other hand types
- Often not possible:
  - One hand may be required for other tasks
  - Thumb-only usage sometimes difficult

Bimanual – MicroRolls

Image source (FU): https://dl.acm.org/citation.cfm?id=1518843

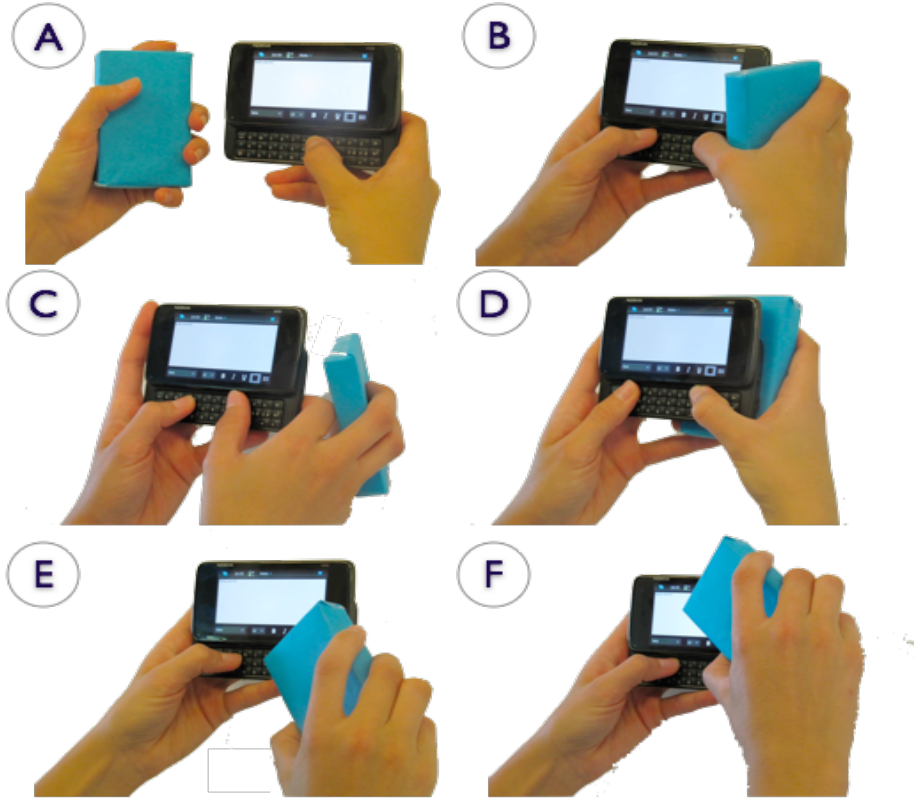
- Thumb-only gestures
- Characterized by rolling without sliding
- 4 linear, 2 circular gestures
  → 6 extra commands
  reachable with single hand

"MicroRolls: expanding touch-screen input vocabulary by distinguishing rolls vs. slides of the thumb", Roudaut et al., CHI 2009

Bimanual – Context

Image source (FU): https://dl.acm.org/citation.cfm?id=1979402

- People use many different strategies to deal with uni-manual context
- E.g. use interfering object as tool (E)
- Highly dependent on usage context (e.g. not possible with coffee mug)



"Ease of juggling: studying the effects of manual multitasking", Oulasvirta et al., CHI 2011

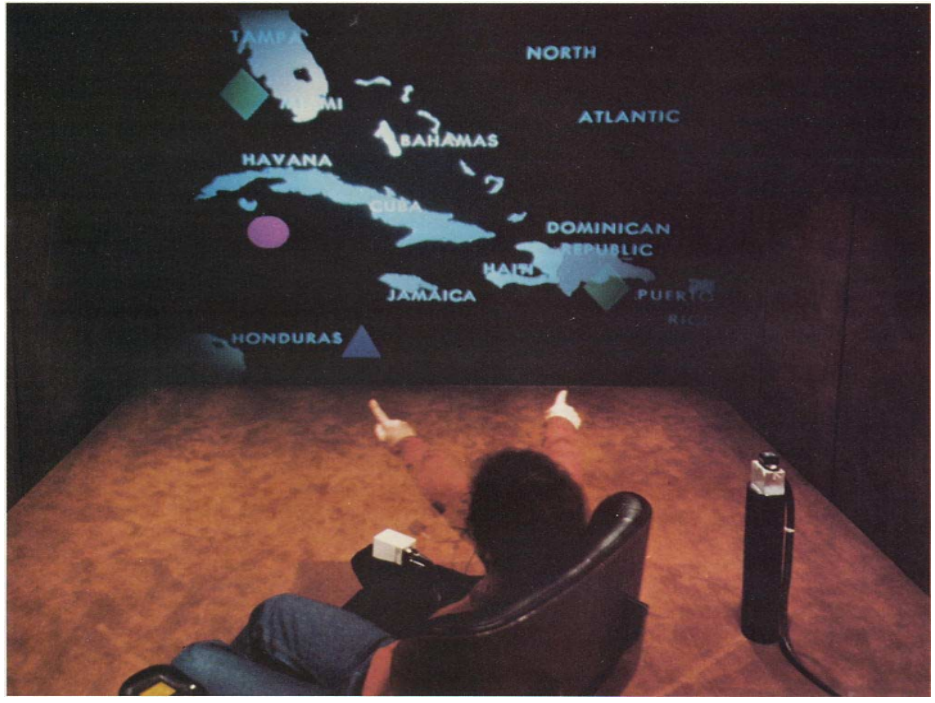
I/O issues: speech (recap)

- Speech input
  - Mostly used for hands-free dialing (in car)
  - Siri, Alexa, Google Now: more complex speech recognition offloaded to cloud service
  - Apparently not widely used (have you ever seen someone talk to Siri like in the commercial?)
- Speech output
  - Mostly used for car navigation
  - Again, not widely used otherwise
- Cultural differences (e.g. US vs. Europe)?

Speech + x – multimodal interaction?

Image source (FU): https://dl.acm.org/citation.cfm?id=800250.807503

- Classic example: “Put-that-there” (1980)
- Speech + pointing
- For mobile devices:
- Context issues: noise environment?

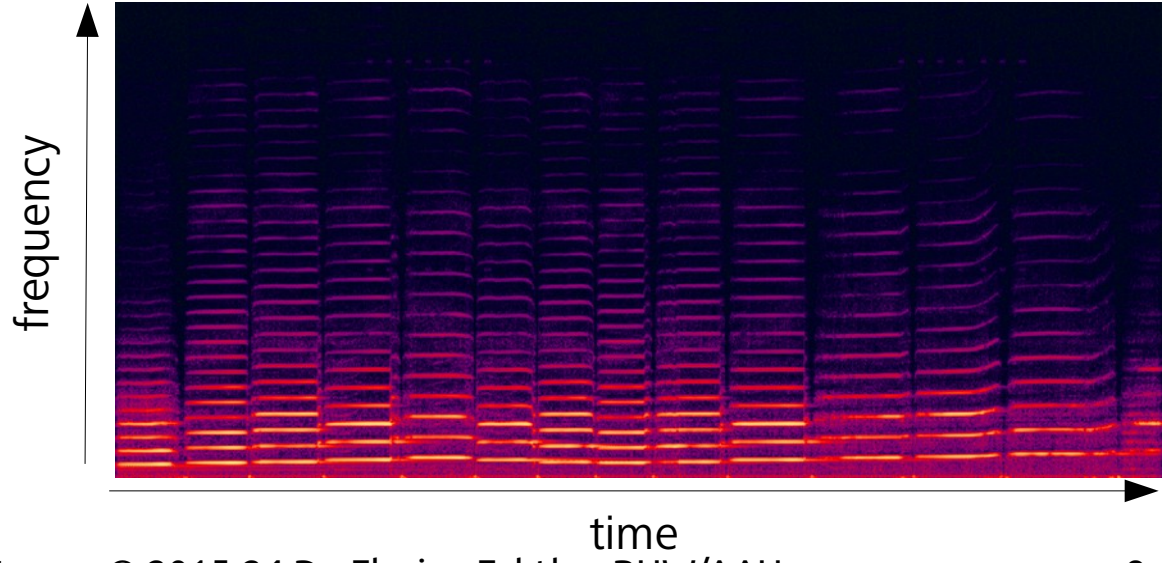


“'Put-that-there': Voice and gesture at the graphics interface”, Bolt, R., SIGGRAPH 1980

Generic sound input (*not* speech)

Image source (CC): https://en.wikipedia.org/wiki/File:Spectrogram_of_violin.png

- e.g. whistling (“keyfinder”)
- Music recording (Shazam etc.)
- Generally based on *fast Fourier transform* (FFT)
- Creates “fingerprint” of audio data, compares with database



I/O issues: vision (recap)

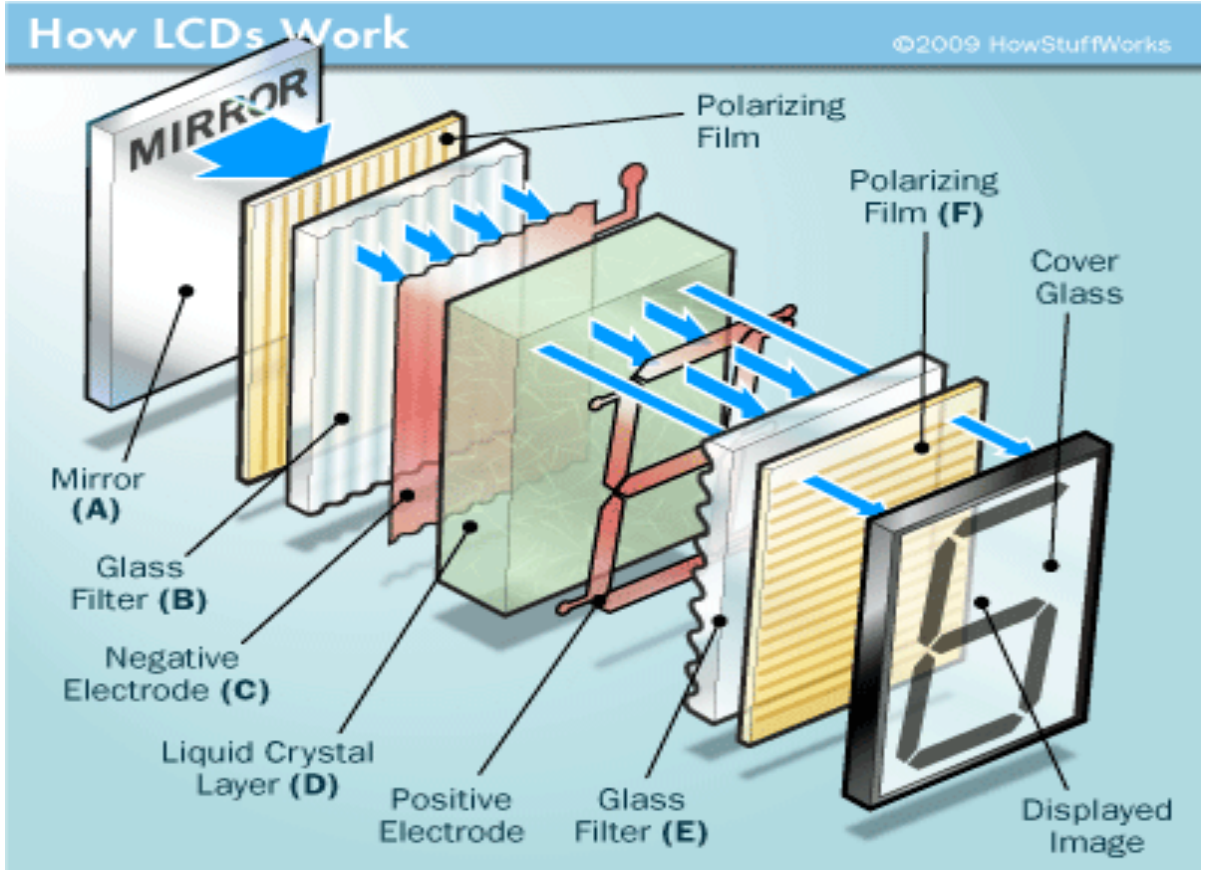
- Visual input (camera)
  - Input of barcodes/QR codes, text recognition (OCR), 3D structure reconstruction (SLAM)
  - Computer vision needs to deal with wildly different lighting conditions (indoor/outdoor)
- Visual output: display
  - Size/resolution: very high information density, suitable information visualization required
  - Brightness/contrast: readable in sunlight?
- Combination: augmented reality

Visual output

- Preface: technology
  - LCD/AMOLED screens
  - E-Ink, 3D displays
- Information visualization

Visual output – Screens: LCD (1)

Image source (FU): http://electronics.howstuffworks.com/lcd2.htm



Visual output – Screens: LCD (2)

- LCD = Liquid Crystal Display
  - Requires constant power to operate
  - Monochromatic on its own → uses color filters
- Usually has own backlight across whole screen
  - Backlight competes with ambient light
    → bad sunlight readability
  - Display absorbs part of white light from backlight
  - Alternative: *transflective* LCD (can use environment light, but more complex/expensive → rarely used)

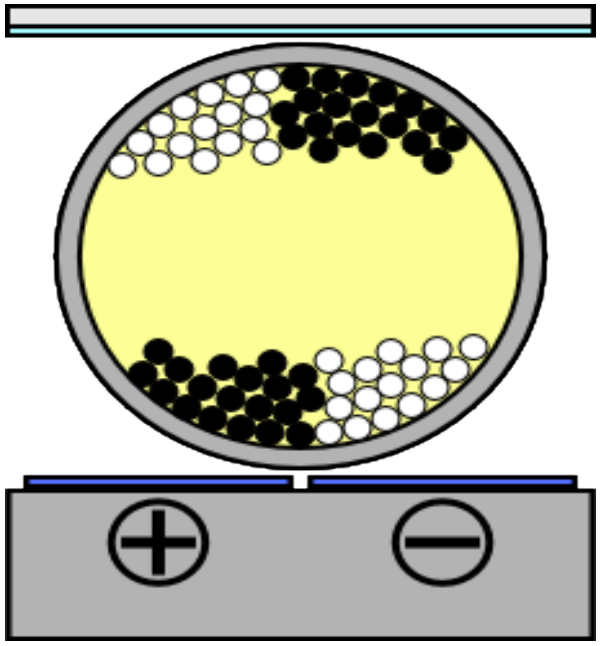
Visual output – Screens: AM(O)LED

- LCDs slowly replaced by AM(O)LED displays
  - Active Matrix (Organic) Light Emitting Diodes
  - Every pixel is a small light source on its own
  - Long-term lifetime issues
- Generally more efficient than LCDs
  - Only generates light which is actually required
  - Lower total brightness (currently)

Visual output – E-Ink displays

Image source (CC): https://commons.wikimedia.org/...Side_view_of_Electrophoretic_display%29_in_svg.svg

- Only particles on the top are visible
- Electric field moves white/black particles to the top
- No energy needed to keep images, only to change (bi-stable)
- High contrast, sunlight readable
- Drawback: too slow for video etc.



Visual output – 3D displays

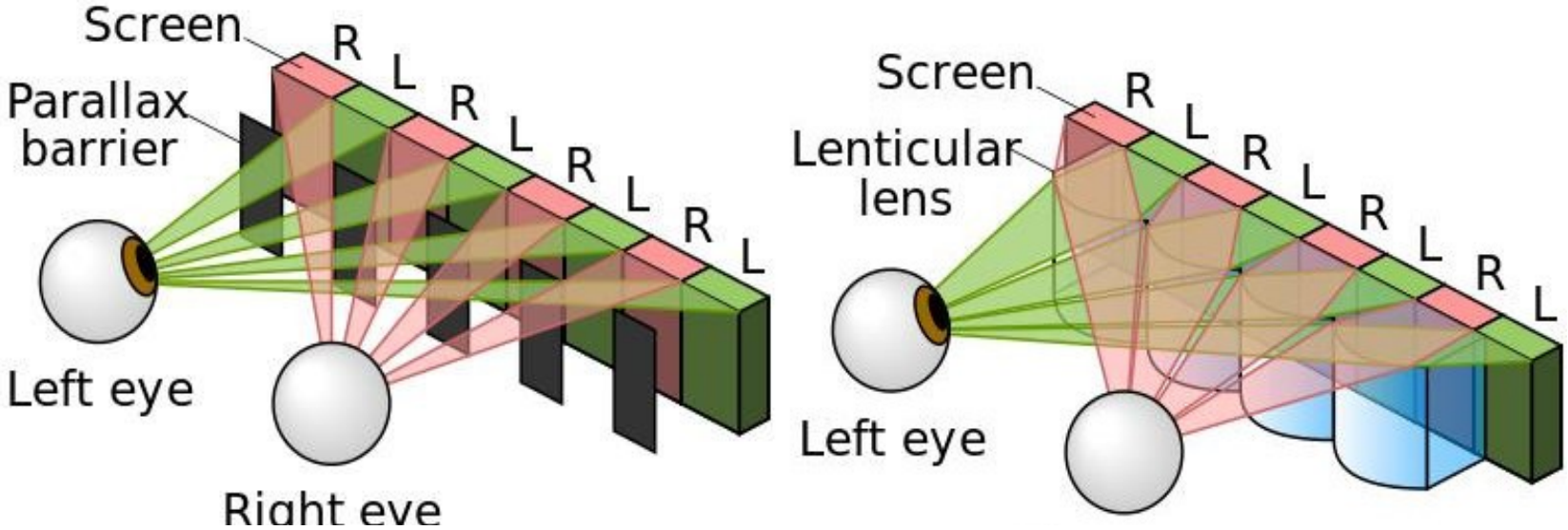
- Deliver two different images to left/right eye
- Problem: multiple depth cues in human vision
  - Stereo disparity (different left/right images) → ok
  - Parallax (different shift due to head motion) → ok
  - Focus depth (different eye accomodation) → fails
  - Body motion (move view by walking) → fails

→ discrepancies can lead to headache, nausea, …

Visual output – Autostereoscopic

Image source (CC): https://en.wikipedia.org/.../File:Parallax_barrier_vs_lenticular_screen.svg

- Parallax barrier / lenticular array
- Only reliable for single, stationary viewer
- Possible extensions: multi-view, user tracking



Visual output – HMDs

Image source (CC): https://en.wikipedia.org/.../File:Oculus_Rift_-_Developer_Version_-_Front.jpg

- HMDs = Head-Mounted Displays
  - e.g. Oculus Rift, Google Daydream, Meta Quest, …
- Small revolution:
  - previous HMDs used 2 small, expensive high-resolution screens & complex optics
  - Now: use simple optics, one large(r) screen + visual correction in software
  - Also possible to use smartphone as display/ sensor platform

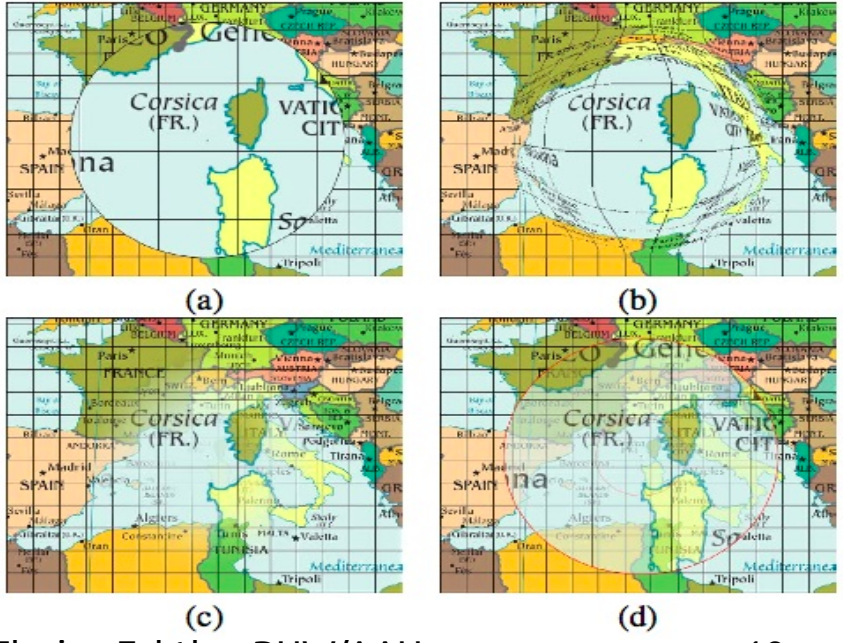
Visual output – InfoVis

- Size/resolution: very high information density
  - Often higher than human eye resolution (“retina screen”)
  - Not sensible to use desktop interface methods (even if similar resolution available)
- Use *Information Visualization* techniques
  - Dedicated discipline with own lecture
  - Lack of screen space is a known problem
  - Example solution: *Focus & Context* methods

Visual output – Focus & Context

Image source (FU): https://dl.acm.org/citation.cfm?doid=1357054.1357264

- Basic idea: show a detailed view (Focus) + overview (Context) at the same time
- Many different variants possible:
  - Context as “World-In-Miniature” (WIM, below)
  - Focus via “lenses” (right)



Visual output – Halo & Wedge

Image source (FU): http://patrickbaudisch.com/projects/halo/, http://patrickbaudisch.com/projects/wedge/

Visual output – Halo & Wedge

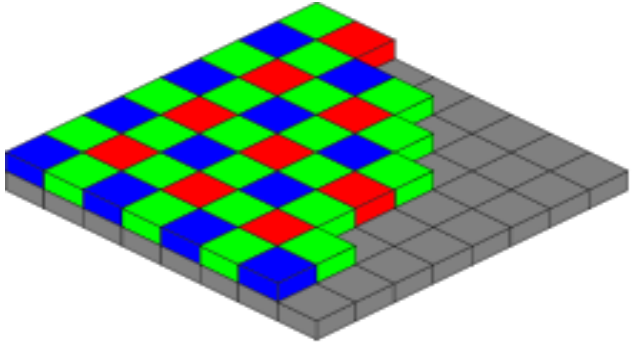
Image source (FU): http://patrickbaudisch.com/projects/halo/

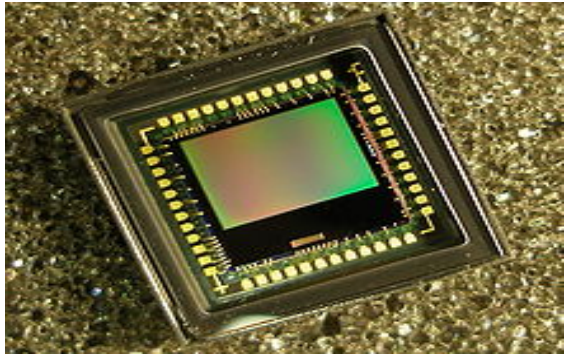
- Projects from 2004 (Halo) & 2008 (Wedge)
- Goal: visualize off-screen locations
  - Focus: map view
  - Context: screen border + halos/wedges
- User doesn't have to change zoom level
- http://patrickbaudisch.com/projects/halo/

Visual input – camera

Image sources (CC): https://wikipedia.org/Charge-coupled_device, https://wikipedia.org/Active_pixel_sensor

- Hardware (right)
  - CCD – higher sensitivity, more expensive
  - CMOS/APS – cheaper, no global shutter (center)
- Only greyscale, uses Bayer color filter (left)
  - 2 x green pixels due to human eye sensitivity





Visual input – camera: barcodes

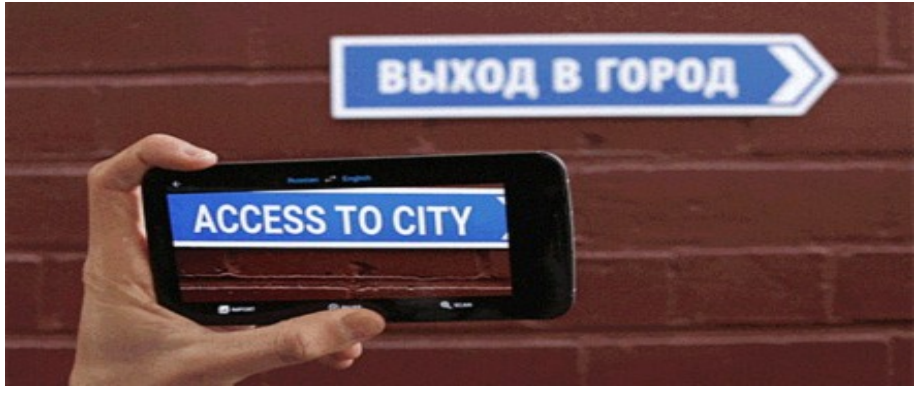
Image source (CC): https://en.wikipedia.org/.../File:Scanning_QR_codes_on_business_cards.jpg

- Two common types of barcodes:
  - EAN product codes (linear, little data, ~ 15 chars)
  - QR codes (2D, up to 3 kB, error correction)
- (Relatively) simple computer vision problem
- “Barcode Scanner” app
  - Can be integrated into own apps as module

Visual input – camera: OCR

Image source (FU): http://cdn1.theweek.co.uk/.../public/9/26/150116-google-translate_0.jpg

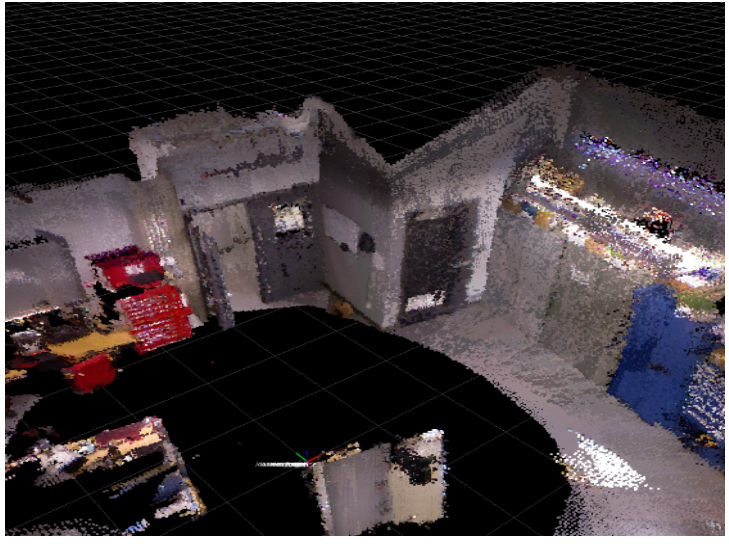
- OCR = Optical Character Recognition
- Complex CV problem (different fonts etc.)
- Can be used to …
  - Scan & *index* documents (just scanning doesn't require OCR)
  - Translate foreign languages (cf. Augmented Reality)



Visual input – camera: SLAM

Image source (FU): https://code.google.com/p/rtabmap/wiki/SLAMDemo

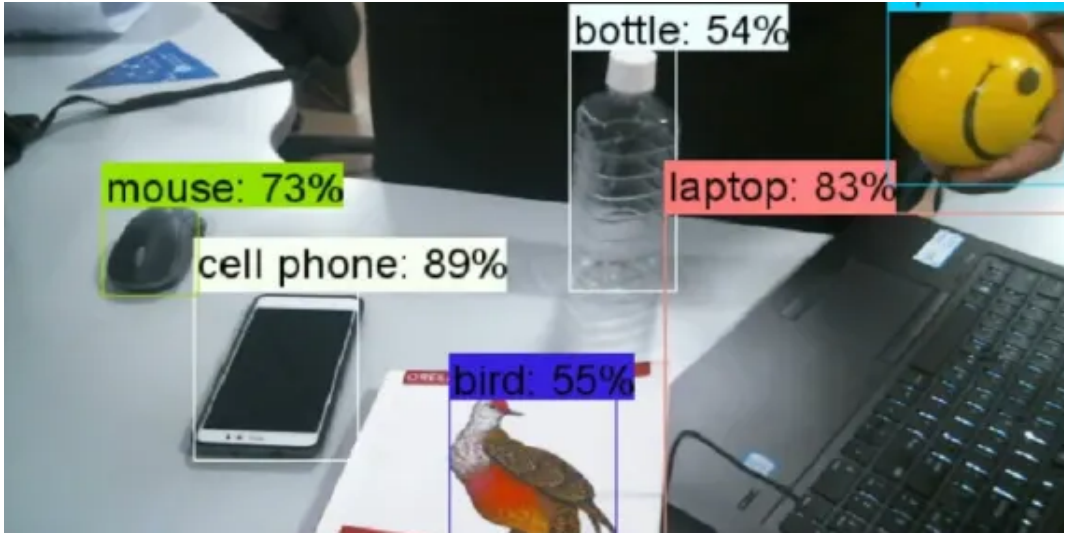
- SLAM = Simultaneous Location And Mapping
  - Creates 3D environment map on-the-fly …
  - … and locates device relative to map
- Usage scenarios:
  - “Inside-out”: device moves within environment → map
  - “Outside-in”: devices moves around object → 3D scan



Visual input – camera: object detection

Image source (FU): https://www.youtube.com/watch?v=MoMjIwGSFVQ

- Usually based on CNN (Convolutional Neural Network)
- Training/learning phase: offline on GPU cluster with large dataset, 10 000+ images with ground-truth labels
- Inference phase: on mobile device, also GPU/TPU accelerated
- CNN needs to be optimized for mobile (cf. Tensorflow Lite)



Visual I/O – Touch Projector

Image source (FU): https://dl.acm.org/citation.cfm?id=1753326.1753671

- All-in-one approach:
  - Multi-screen visual output
  - Visual + touch input
- Touches on mobile device are “projected” onto larger screen

“Touch projector: mobile interaction through video”, Boring et al., CHI 2010

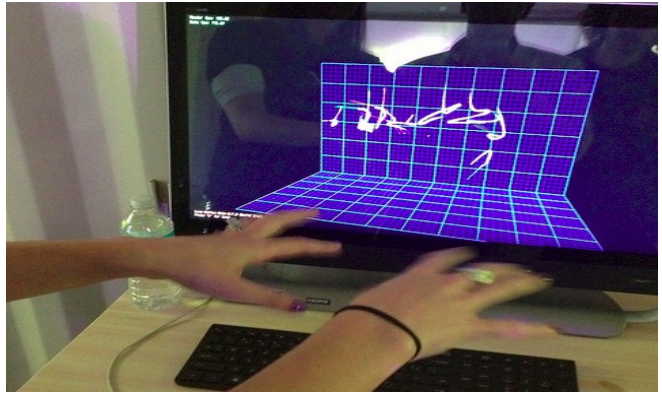
I/O issues: other channels

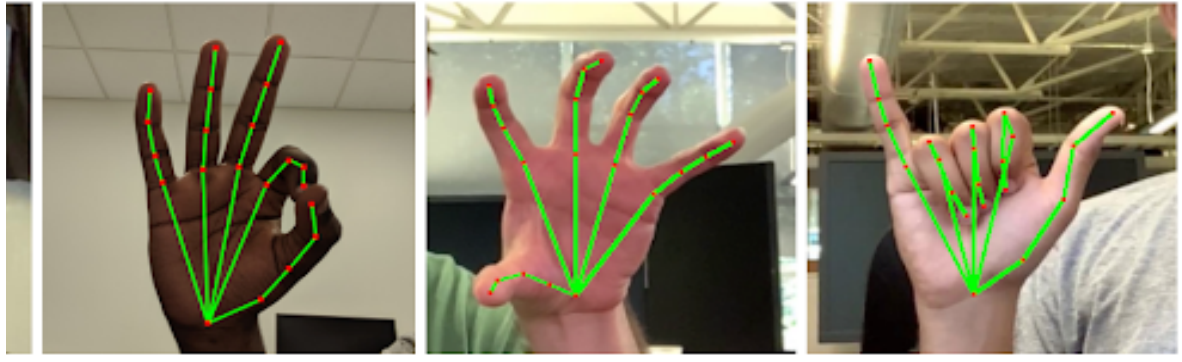
- Bio sensors
  - Fingerprint, heart rate, skin conductivity
  - Privacy issues?
- Spoilt for choice? Too “exotic” for user?

Other channels – Hand/Body Stance

Image source (CC): https://www.flickr.com/photos/davidberkowitz/8598269932/
Image source (FU): https://ai.googleblog.com/2019/08/on-device-real-time-hand-tracking-with.html

- Complex model of human skeleton required
  - Computationally intensive
  - Needs external sensors (e.g. Leap Motion – left) or CNN (right)

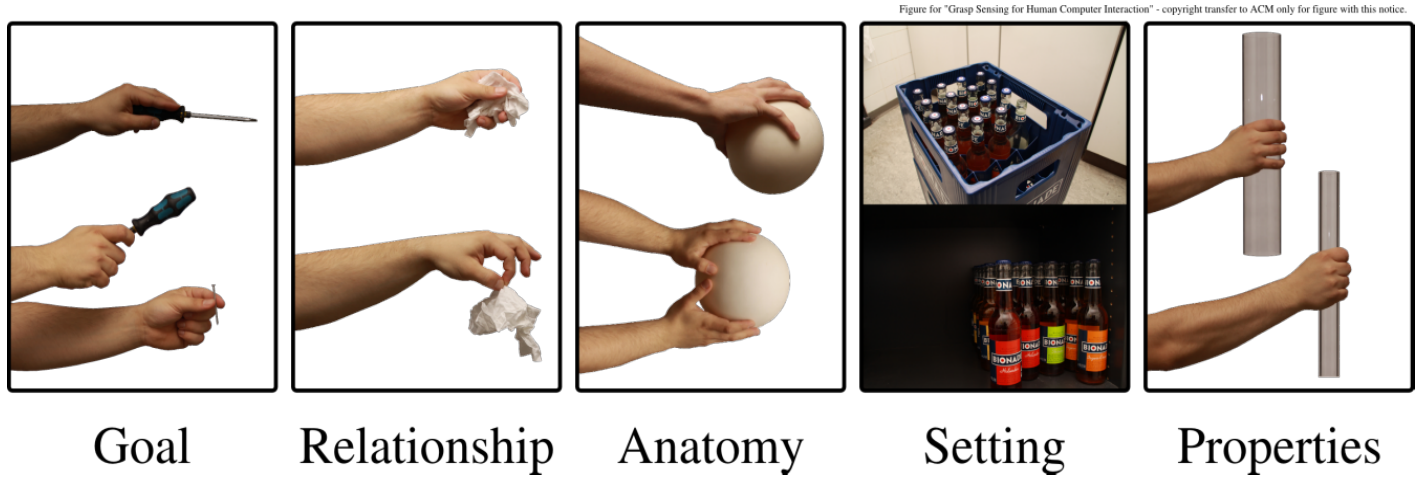




Other channels – Grasp

Image source (FU): https://dl.acm.org/citation.cfm?id=1935701.1935745

- Alternative to full body stance: grasp shape
  - Can use *on-device* sensing (capacitive/optical)
  - Enables conclusions about many facets of context

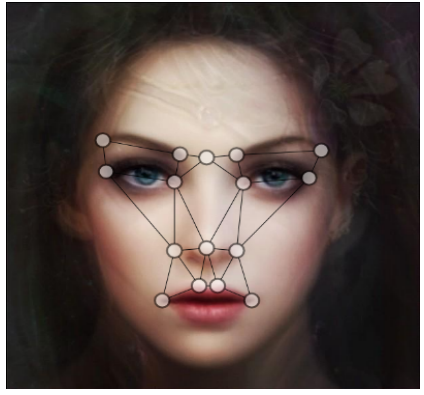


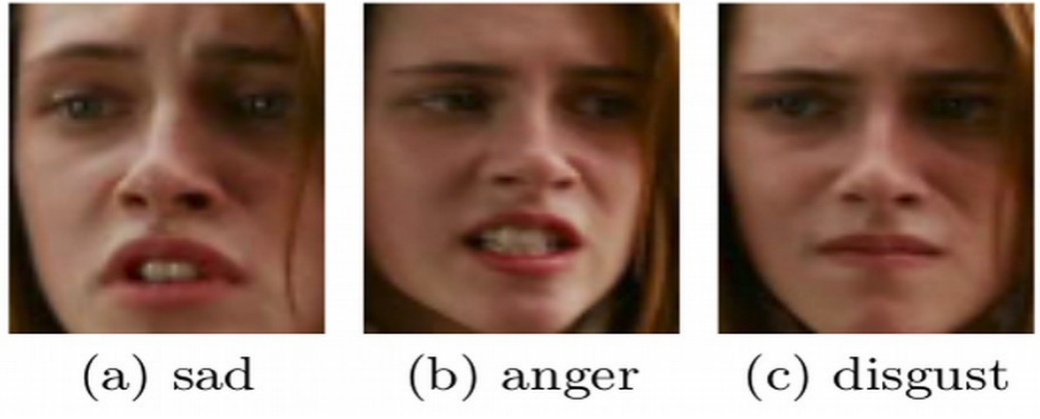
"Grasp sensing for human-computer interaction", Wimmer et al., TEI 2011

Other channels – facial expression

Image source (CC): https://www.flickr.com/photos/averageface/8260432583/
Image source (FU): https://cvhci.anthropomatik.kit.edu/publications_924.php

- Active area of research in computer vision
  - Still unreliable for multiple different persons
  - Context issues: lighting, motion?
- Usage: mood (= context) detection, security





Other channels – facial expression

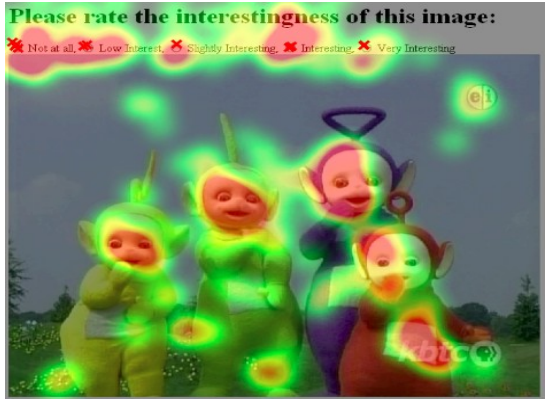
Image source (FU): https://static.giga.de/wp-content/uploads/2017/11/animoji-iphone-x.png

- *Interpretation* of facial expression is difficult (even for humans)
- *Mapping* to an avatar is easier (“Animoji”)
- Usually also solved with CNN to detect feature points in face

Other channels – eye/gaze tracking

Image source (CC): https://www.flickr.com/photos/smieyetracking/14970688193/
Image source (CC): https://www.flickr.com/photos/andyed/387442396/

- Dedicated device or integrated into phone
  - Determines gaze direction of user (i.e. also provides pointing device, attention level)
  - Also: concentration level (pupil dilation, saccade frequency = involuntary eye movements)



Other channels – Heart rate/ECG

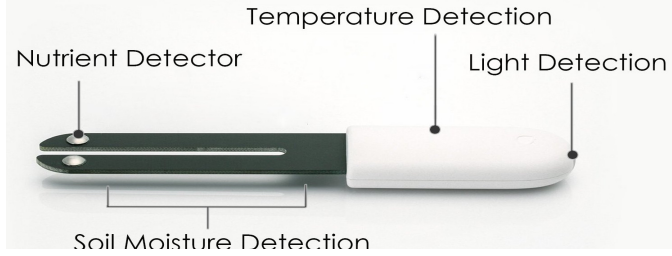
Image source (PD): https://commons.wikimedia.org/wiki/File:Wrist-oximeter.jpg

- Heart rate/blood oxygen saturation can be measured using IR light (pulse oximeter)
- Built into recent smartwatches/fitness bands
- Even possible with smartphone camera! (How?)

Other channels – Environment

Image sources (FU): http://www.exp-tech.de/seeed-studio-grove-gas-sensor-mq2, https://www.amazon.de/dp/B06XKXFFNZ/, https://techcrunch.com/2017/04/30/monit/

- Environmental sensors:
  - Gas concentration (e.g. methane, NO_x, CO_2 etc.)
  - Temperature, moisture, sunlight, …
- Usage scenarios
  - Citizen science projects, e.g. for air quality
  - Plant & diaper monitoring (seriously :-)



Other channels – fingerprint

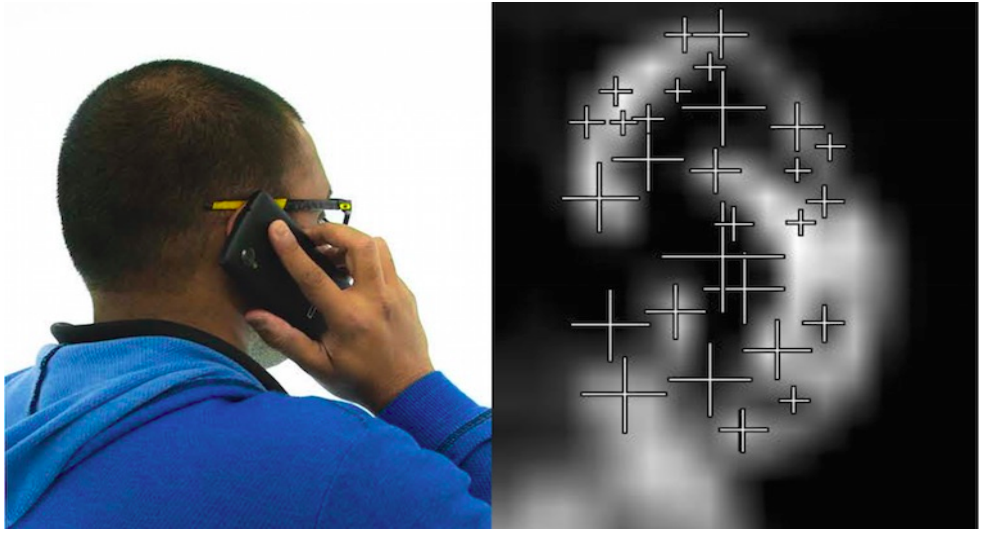
Image source (CC): https://en.wikipedia.org/.../File:Fingerprint_scanner_identification.jpg

- Many recent phones have fingerprint sensors
  - Used for unlocking device, authorizing transactions, …
  - Usually capacitive, sometimes optical
- Could be integrated directly into screen → ?
  - Immediate authentication
  - Finger identification
  - Increased precision

Other channels – body shape

Image source (FU): https://dl.acm.org/citation.cfm?id=2702518

- Detects shape of touching body part
- Uses capacitive touch screen to create rough image of body part (hand, ear, knuckles, …)
- Used for authentication, e.g. only correct ear is able to answer phone



"Bodyprint: Biometric User Identification on Mobile Devices Using the Capacitive Touchscreen to Scan Body Parts", Holz et al., CHI 2015

Other channels – brain currents (1)

- BCI = Brain-Computer Interface
  - Measures extremely low currents created by brain activity
- In theory, the non-plus-ultra of interfaces
  - Hands-free, silent, no visual attention needed, …
- In practice: very difficult to …
  - Set up (electrodes on head)
  - Learn (significant practice required)
  - Use (multiple error sources)

Other channels – brain currents (2)

Image source (CC): https://www.flickr.com/photos/h860a/5827080225/
Image source (FU): http://recherche.parisdescartes.fr/.../Moyens-Techniques/EEG-Platform

- Consumer-grade hardware can currently only determine general states (relaxed, focused, …)
- Medical-grade hardware: better capabilities, but even more complex setup (e.g. MRI)

The End