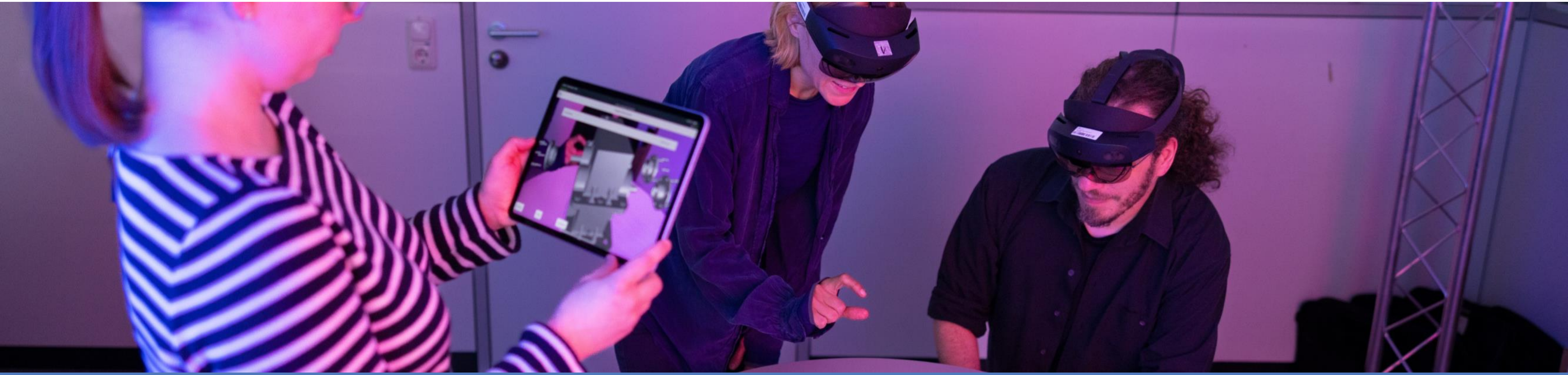
FRANKFURT UNIVERSITY OF APPLIED SCIENCES



Mixed Reality Technologies

Human-Computer Interaction Lecture

Slides adapted from the Game Design Workshop by V. Schwind, Image from: https://pxhere.com/de/photo/764632

Learning Goals

- Understand …
  - Why accessibility is important in HCI
  - That there are legal and usability requirements
  - Different types of disabilities
  - Principles of universal design
- Be able to explain …
  - Universal design principles
  - Soft- and hardware for people with disabilities

What is Mixed Reality?

Mixed Reality

Real Environment	Augmented Reality (AR)	Augmented Virtuality (AV)	Virtual Environment
	 		 

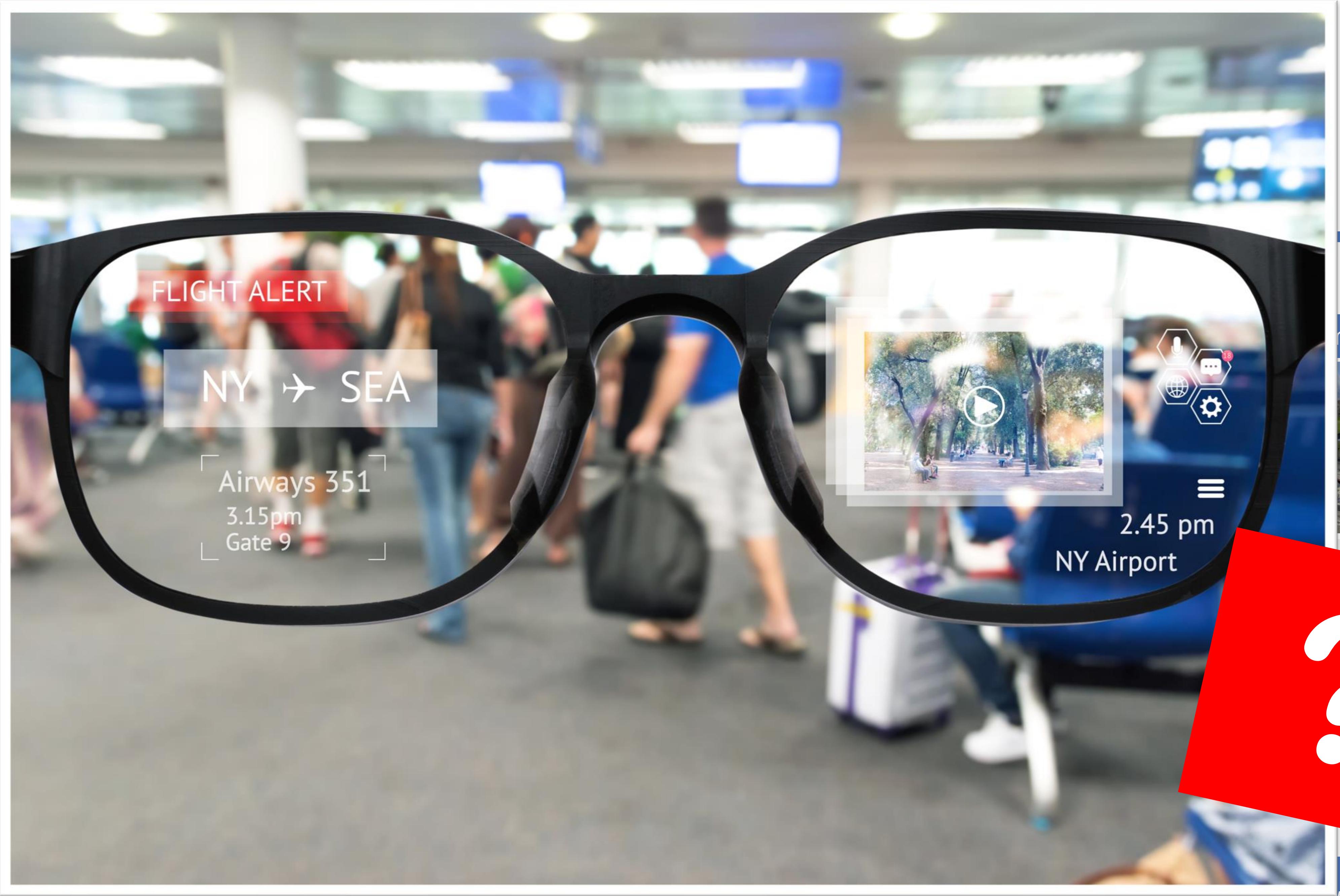
[1] Milgram, Paul; H. Takemura; A. Utsumi; F. Kishino (1994). "Augmented Reality: A class of displays on the reality-virtuality continuum". Proceedings of SPIE - The International Society for Optical Engineering Vol. 2351. Retrieved 2021-06-01.

Images from https://www.mein-erlebnis.blog/ein-tag-in-esslingen/, https://industrieanzeiger.industrie.de/technik/logistik/virtual-reality-cave-zur-materialfluss-und-lagerplanung/ https://www.bechtle.com/ueber-bechtle/newsroom/it-solutions/2022/fuenf-use-cases-mit-der-mixed-reality-brille-hololens-2

Wha

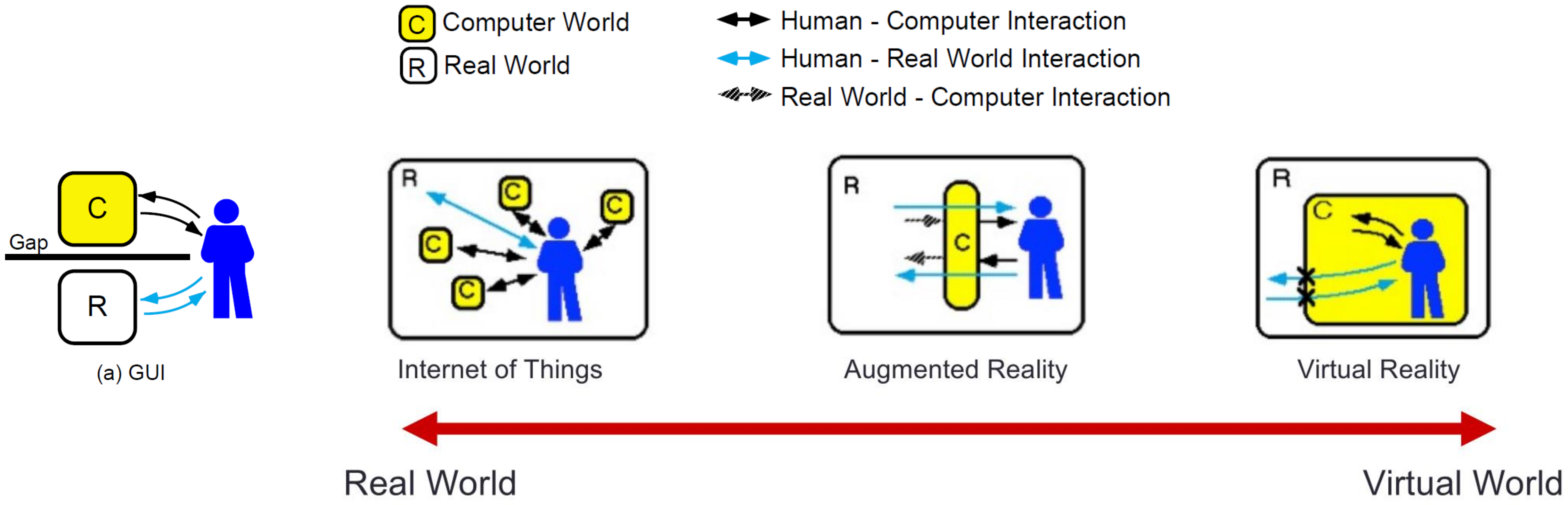
Re

nent



[1] Milgra
Optical E
Images fr
https://ww

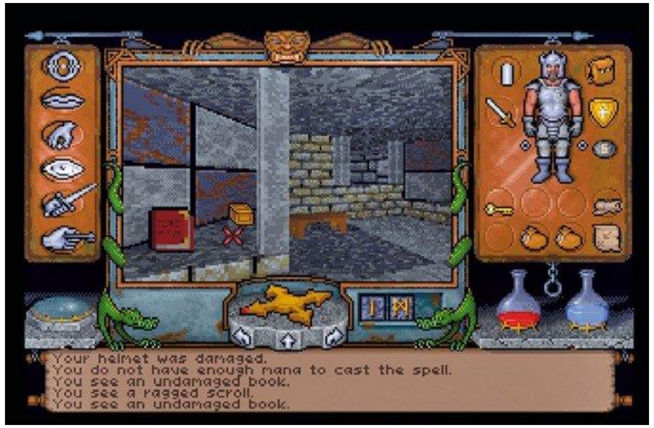
Mediated Interaction



Jun Rekimoto and Katashi Nagao. 1995. The world through the computer: computer augmented interaction with real world environments. In Proceedings of the 8th annual ACM symposium on User interface and software technology (UIST '95).

Mixed Reality Displays

Window on Virtual Environment



Ultima Underworld: The Stygian Abyss (1992)

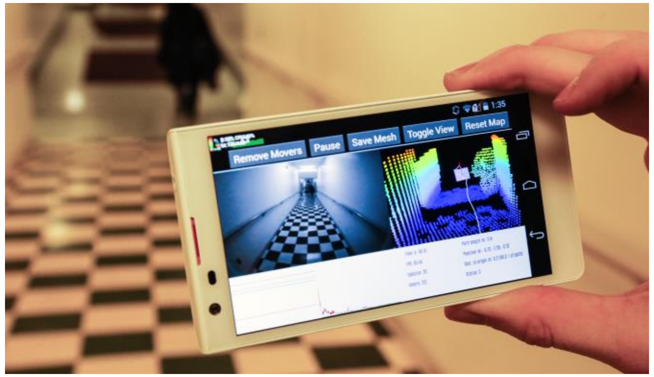
Virtual Environment Cave

CAVE, University of Illinois at Chicago

Virtual Reality Headset

Oculus Rift with Leap Motion

Augmented Reality Display



Google Project Tango (2017)

Augmented Reality Projection

Augmented Projection, Reed, S.(2014)

Augmented Reality Headset

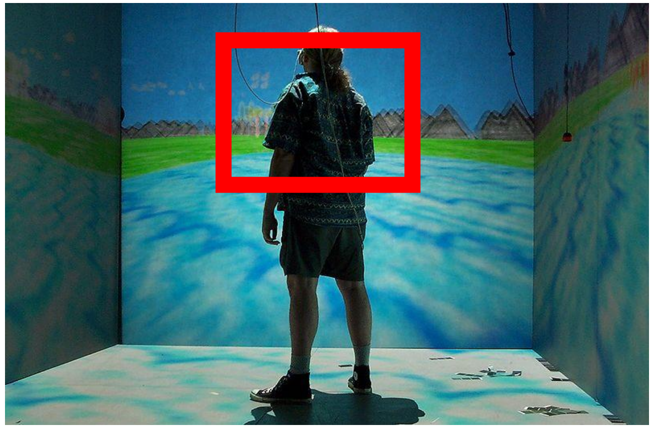
Microsoft Holo Lens (2017)

Mixed Reality Displays

Window on Virtual Environment

Ultima Underworld: The Stygian Abyss (1992)

Virtual Environment Cave

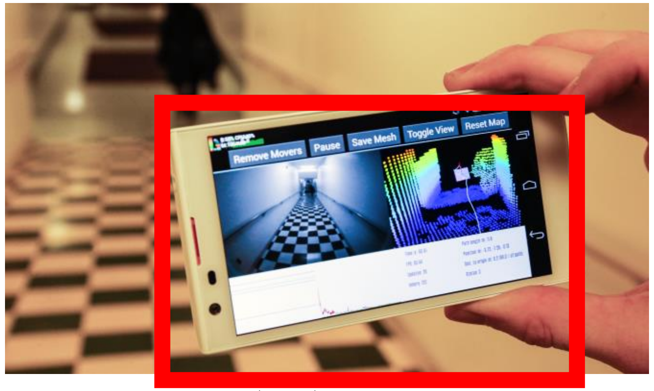


CAVE, University of Illinois at Chicago

Virtual Reality Headset

Oculus Rift with Leap Motion

Augmented Reality Display



Google Project Tango (2017)

Augmented Reality Projection

Augmented Projection, Reed, S.(2014)

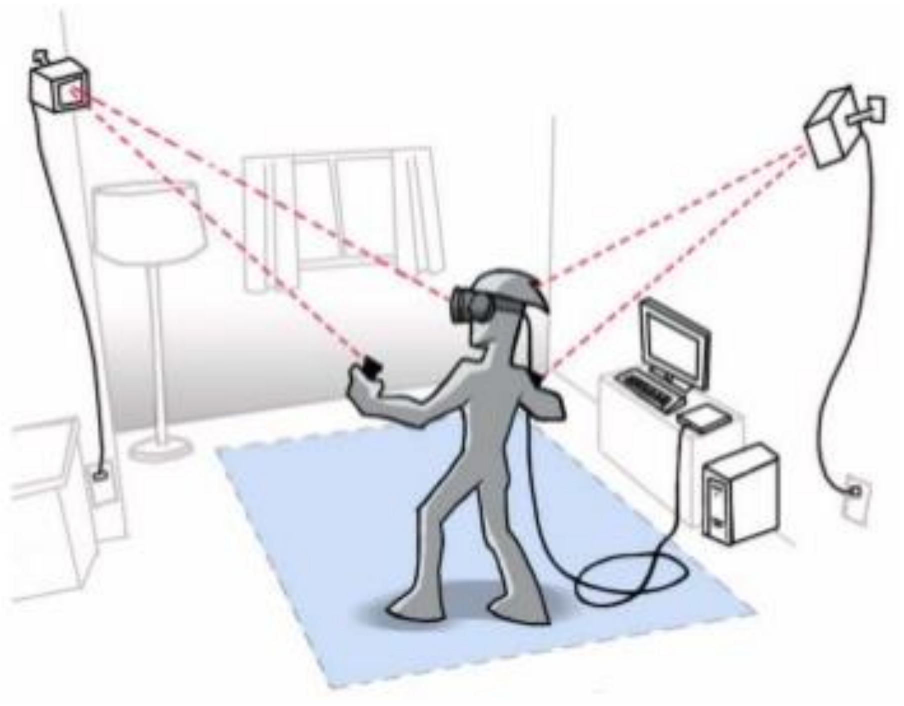
Augmented Reality Headset

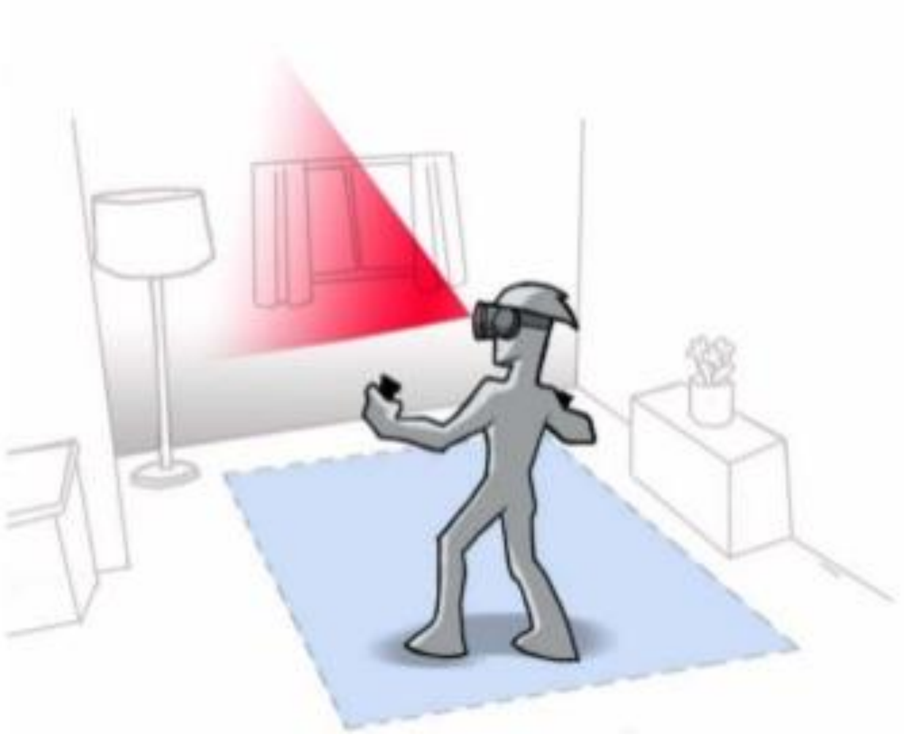
Microsoft Holo Lens (2017)

Mixed Reality Displays

Image: Rufat Rzayev, Susanne Korbely, Milena Maul, Alina Schark, Valentin Schwind, and Niels Henze. 2020. Effects of Position and Alignment of Notifications on AR Glasses during Social Interaction. Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society. Association for Computing Machinery, New York, NY, USA, Article 30, 1–11. DOI: https://doi.org/10.1145/3419249.3420095

Tracking





Outside-In Tracking

External sensors determine position und orientation of the user

Inside-Out Tracking

The device determines the position and orientation of the user

Images: Qualcomm (2017)

Degrees of Freedom (DoF)

3 DoF

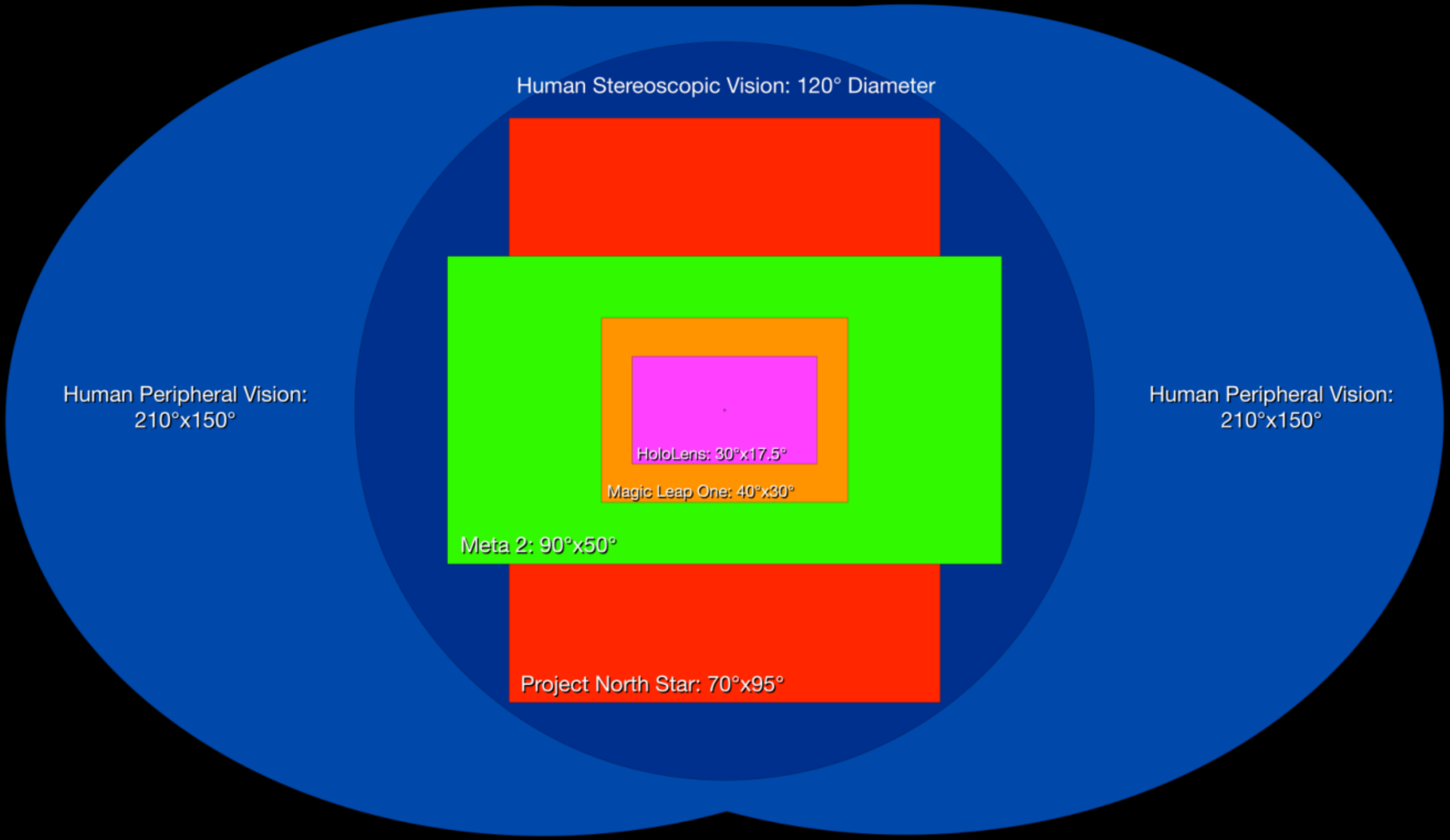
Orientation tracking only
360° watching

6 DoF

Orientation and position tracking in a "room-scale" VR

Images: Qualcomm (2017)

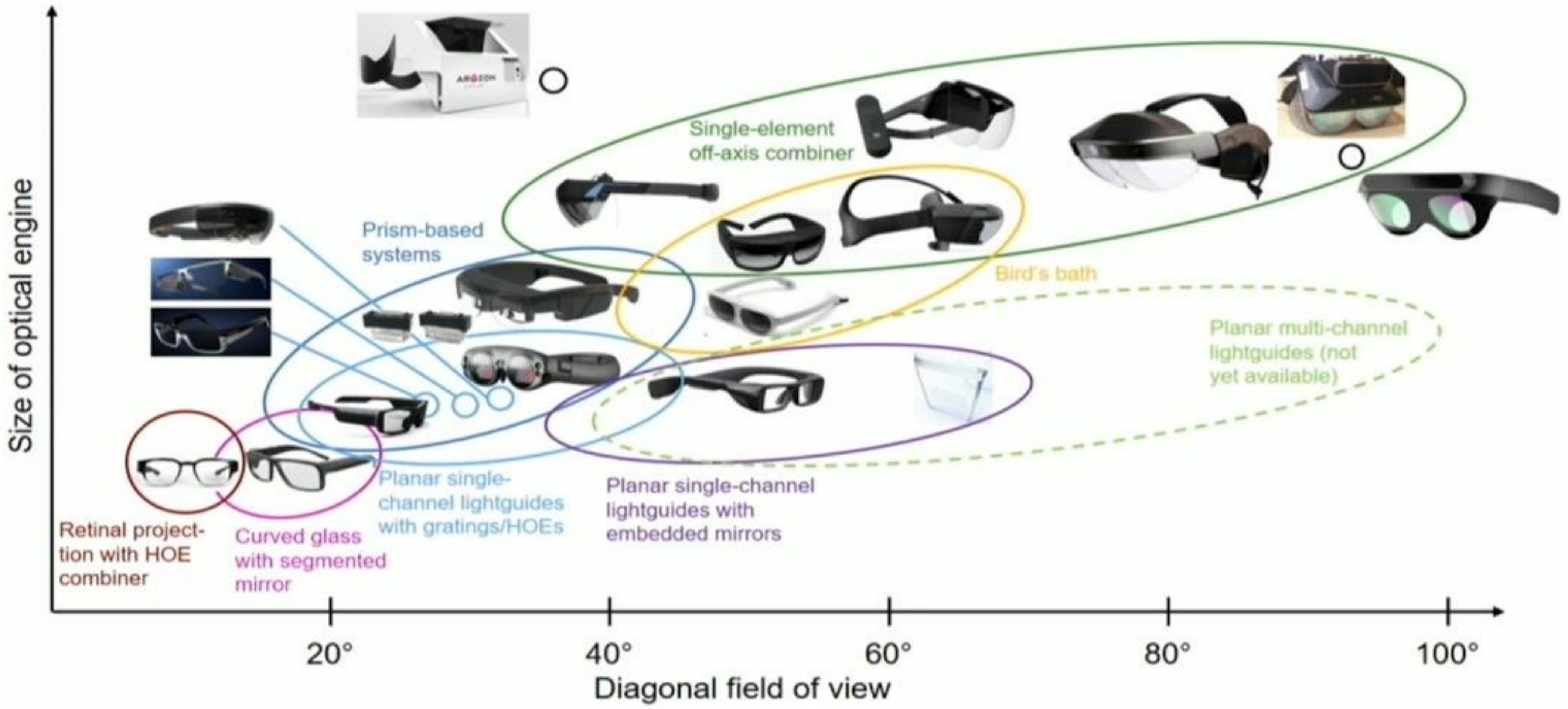
Fields of View in Augmented Reality Head Mounted Displays



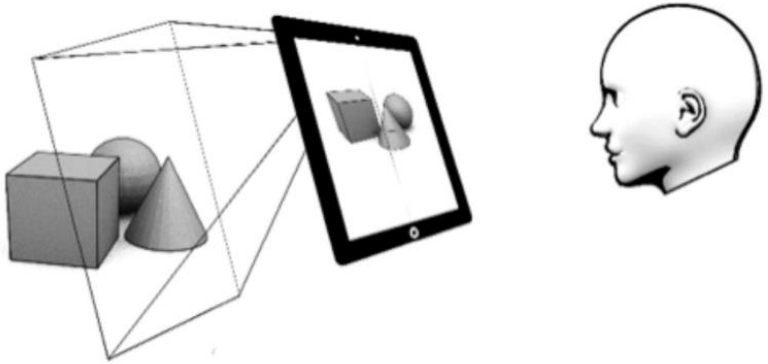
v0.2 by @bfineman

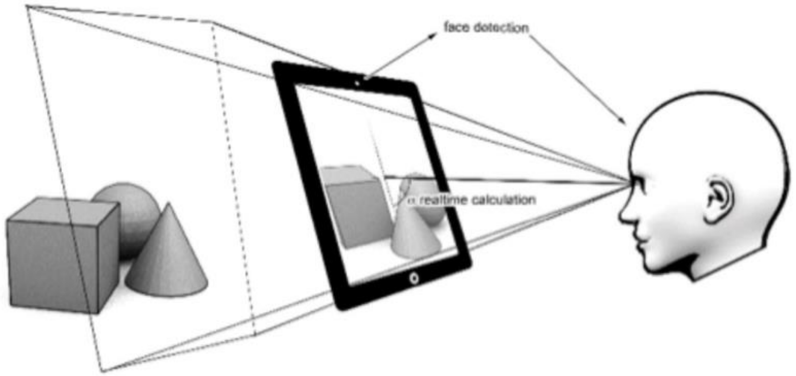
Qualitative categorization of current AR systems according to FOV and optical engine size

tooz technologies



Handheld AR / „The virtual window“



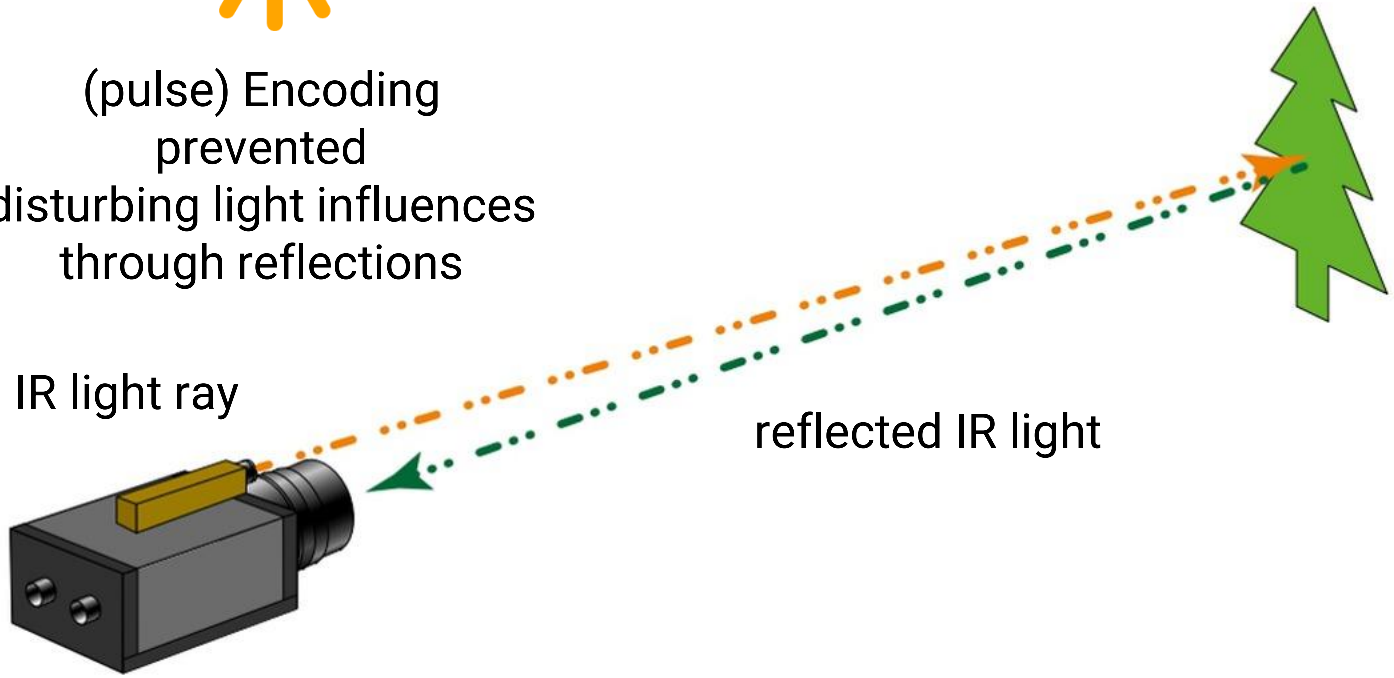



POWER LEVEL
TRAIN
FEED
PLAY
STORE

Time-of-Flight

(pulse) Encoding
prevented
disturbing light influences
through reflections

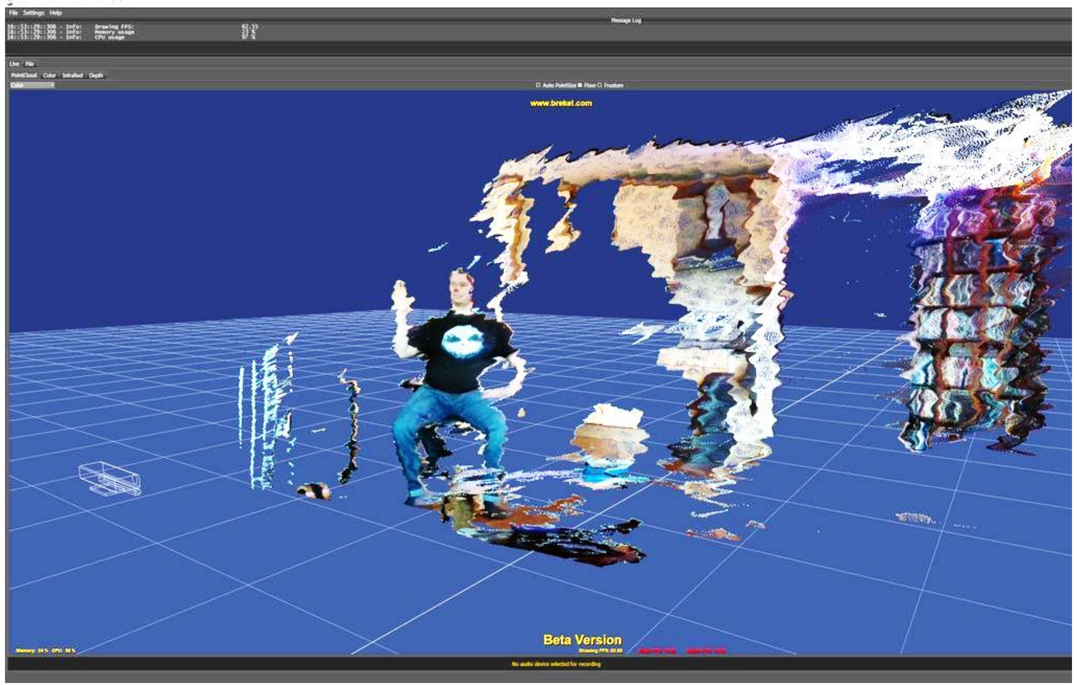
coded IR light ray



reflected IR light

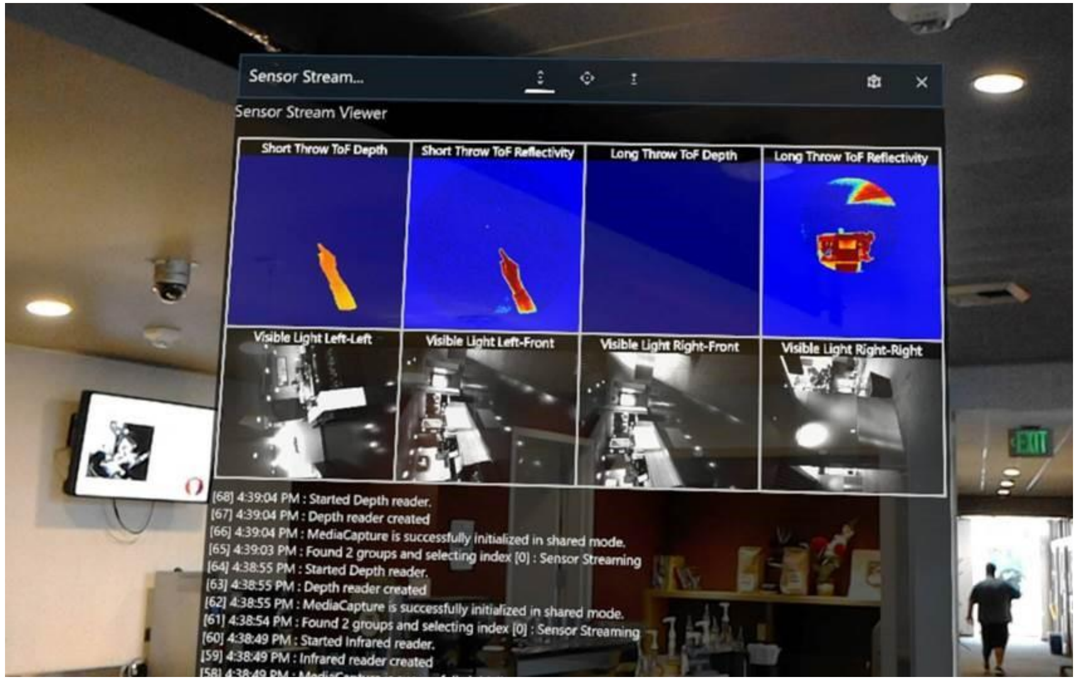
Depth sensor, measures the
travelled time of light (e.g., 100 ns)

$$D = \frac{c_{Luft} * t}{2} = \frac{299.710\frac{km}{s} * 100ns}{2} \sim 15 \text{ m}$$



Outside-In Scene Reconstuction

Time-of-Flight sensor tracking
MS Kinect v.2

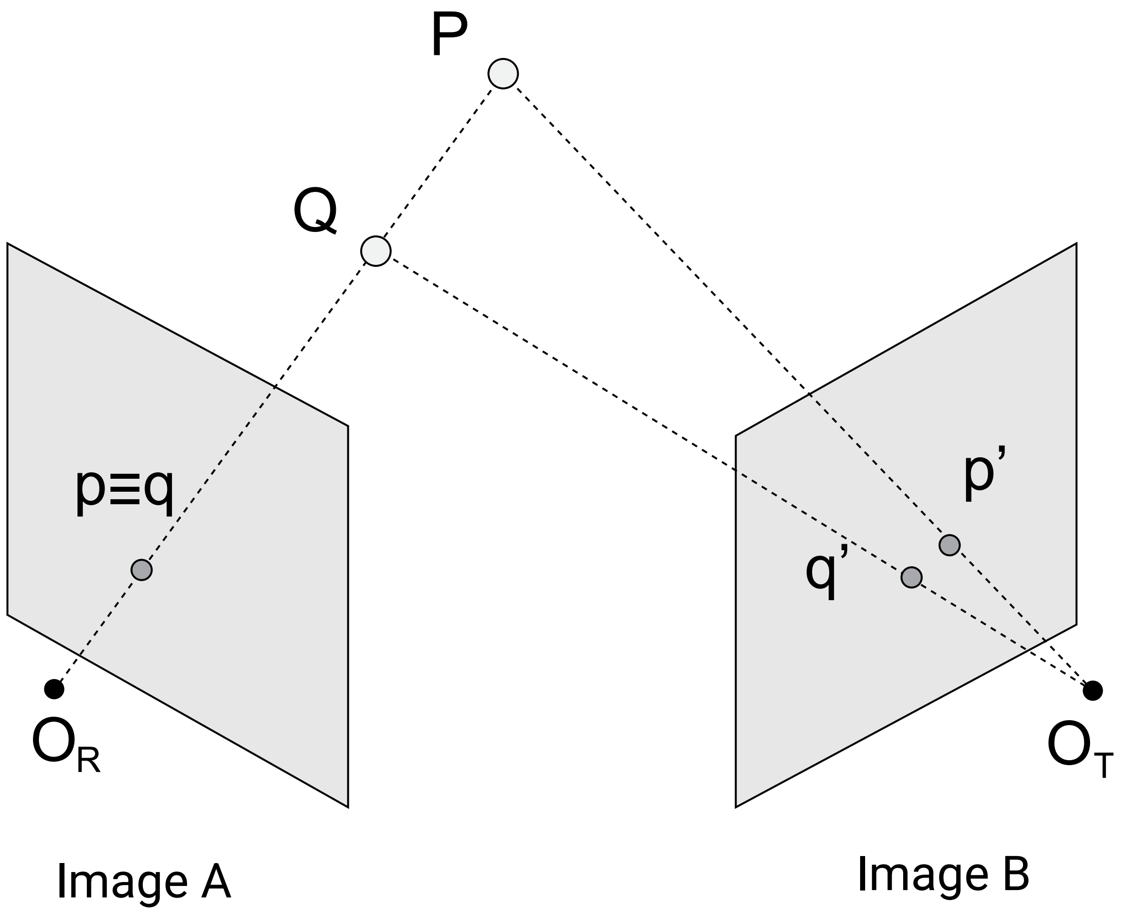


Inside-Out Scene Reconstruction

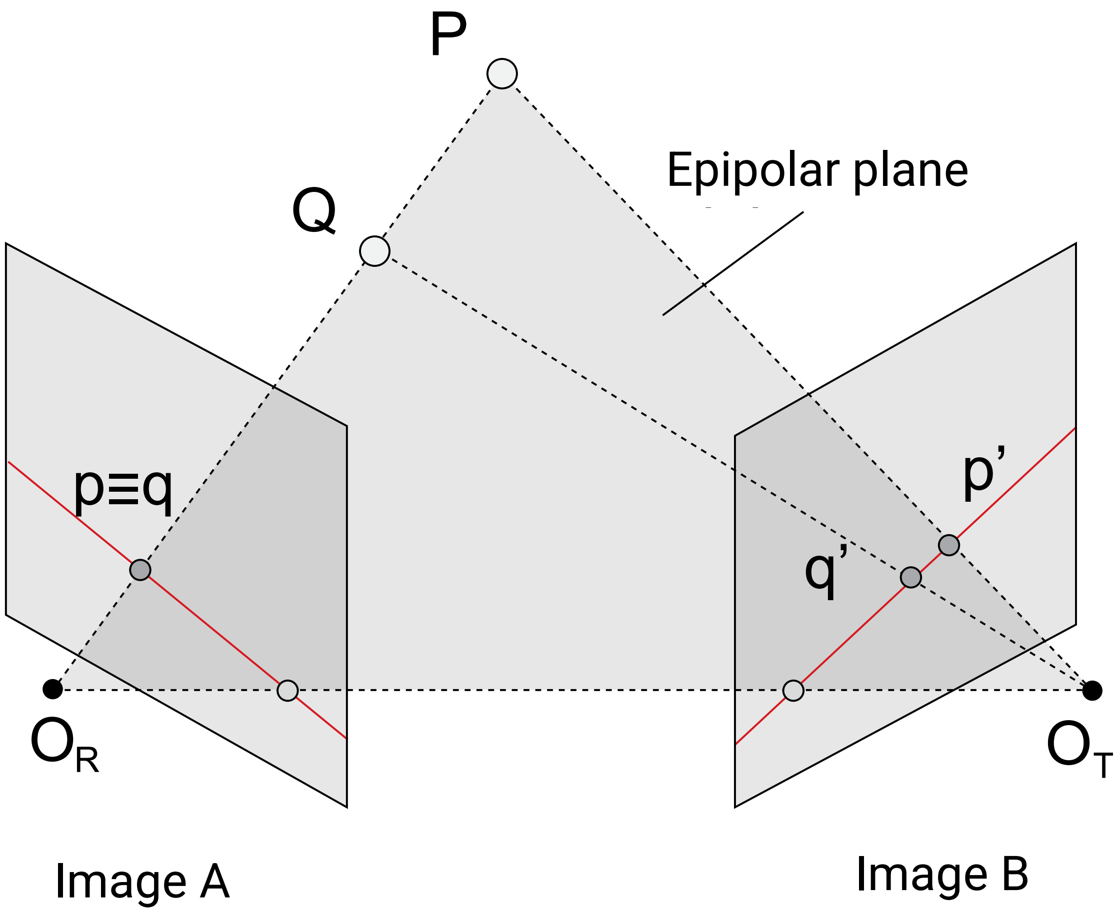
Time-of-Flight sensor tracking
Microsoft HoloLens v2

Stereoscopy

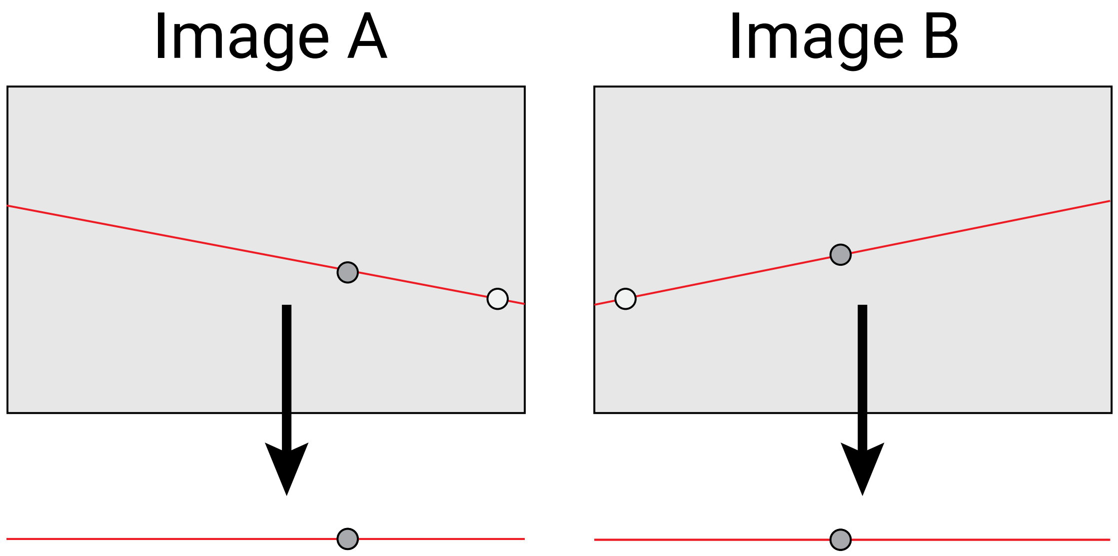
Depth image reconstruction



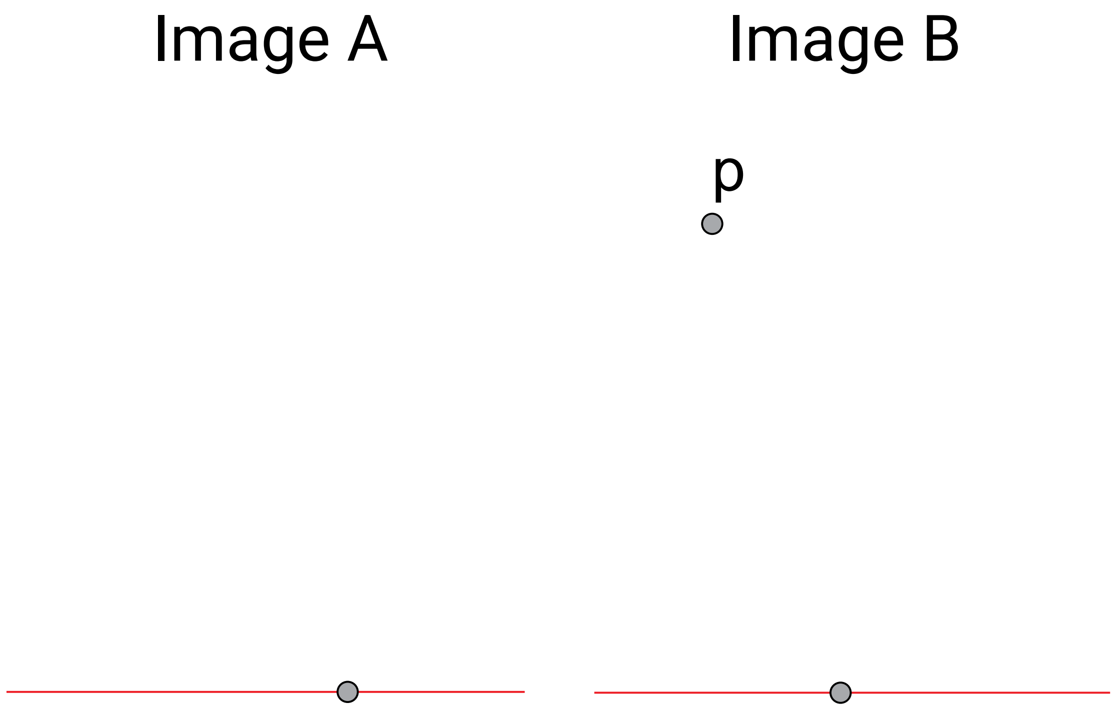
Depth image reconstruction



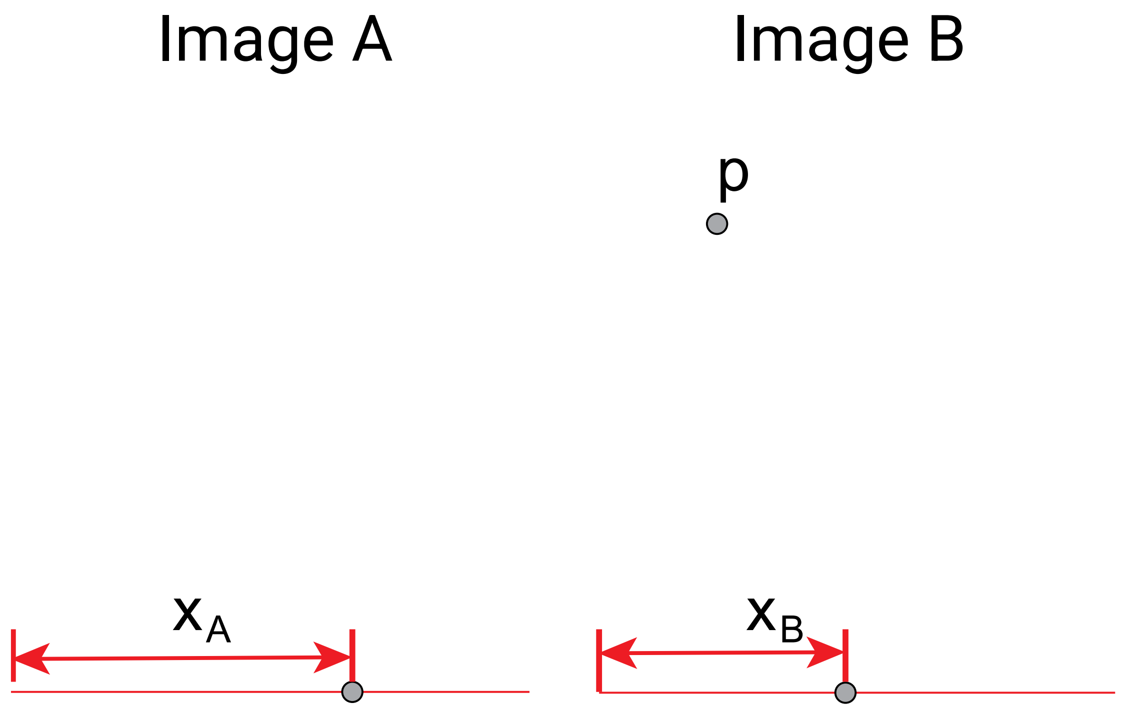
Depth image reconstruction



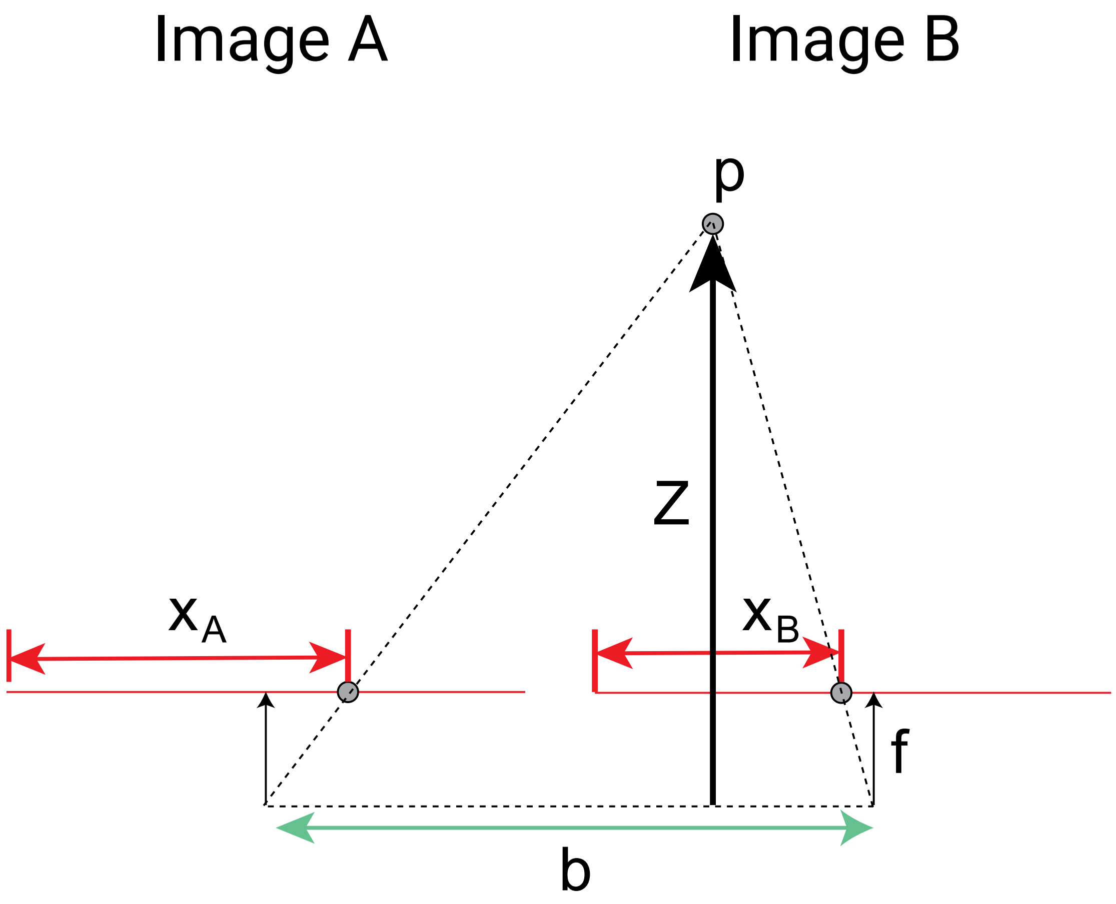
Depth image reconstruction



Depth image reconstruction



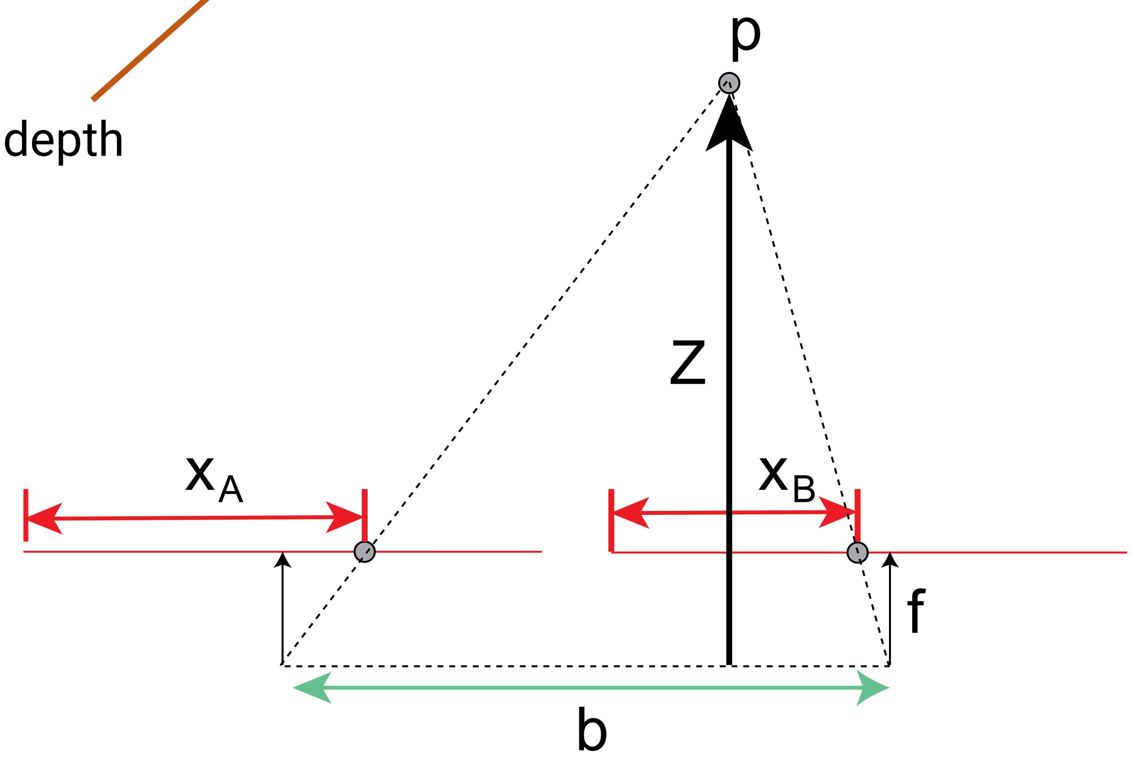
Depth image reconstruction



Depth image reconstruction

$$\frac{b}{Z} = \frac{(b + x_B) - X_A}{Z - f} \quad \rightarrow \quad \boxed{Z} = \frac{bf}{X_A - X_B} = \frac{bf}{d}$$

depth

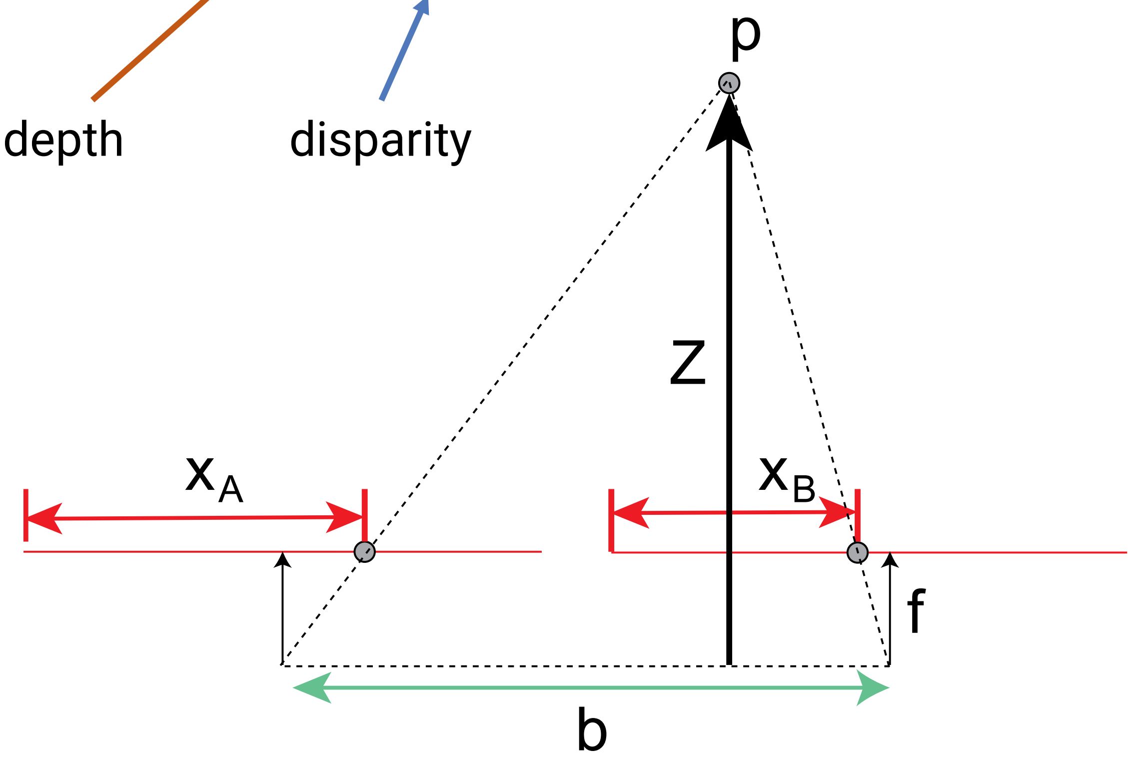


Depth image reconstruction

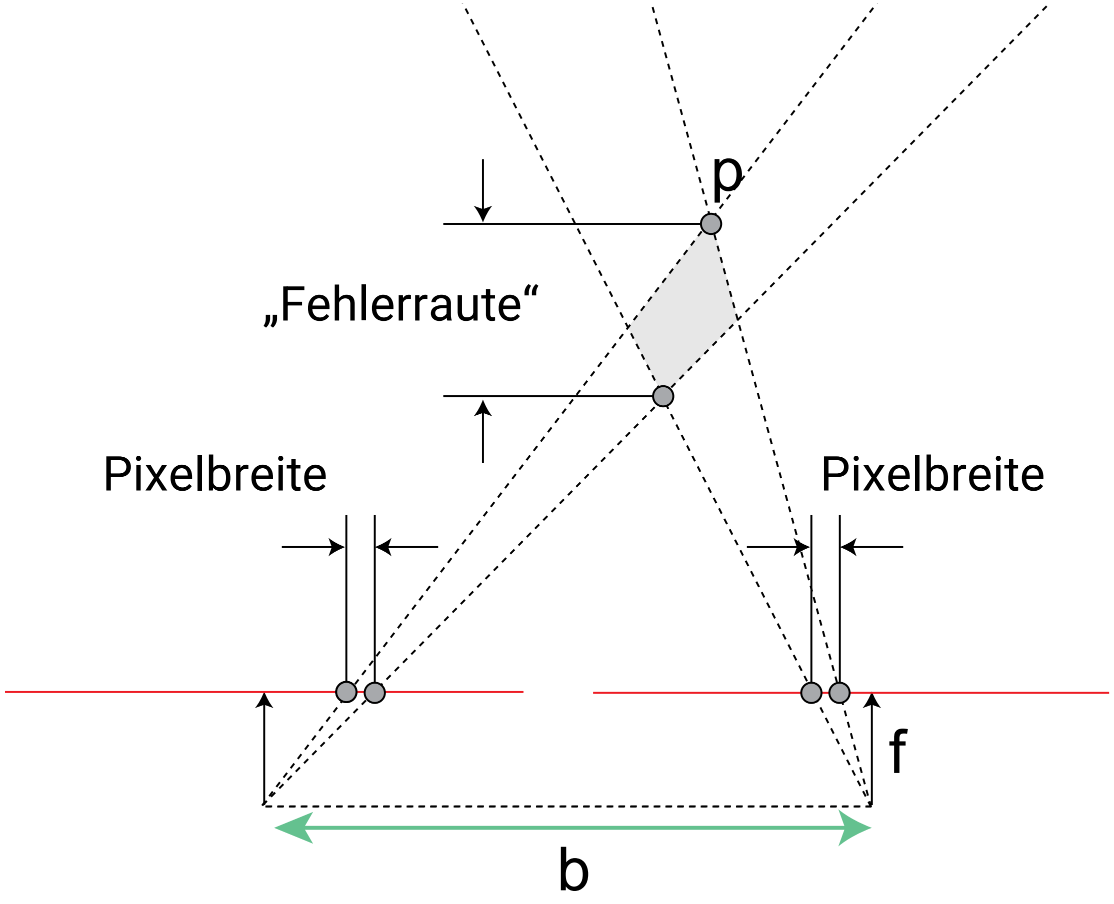
$$\frac{b}{Z} = \frac{(b + x_B) - X_A}{Z - f} \rightarrow Z = \frac{bf}{X_A - X_B} = \frac{bf}{d}$$

depth

disparity

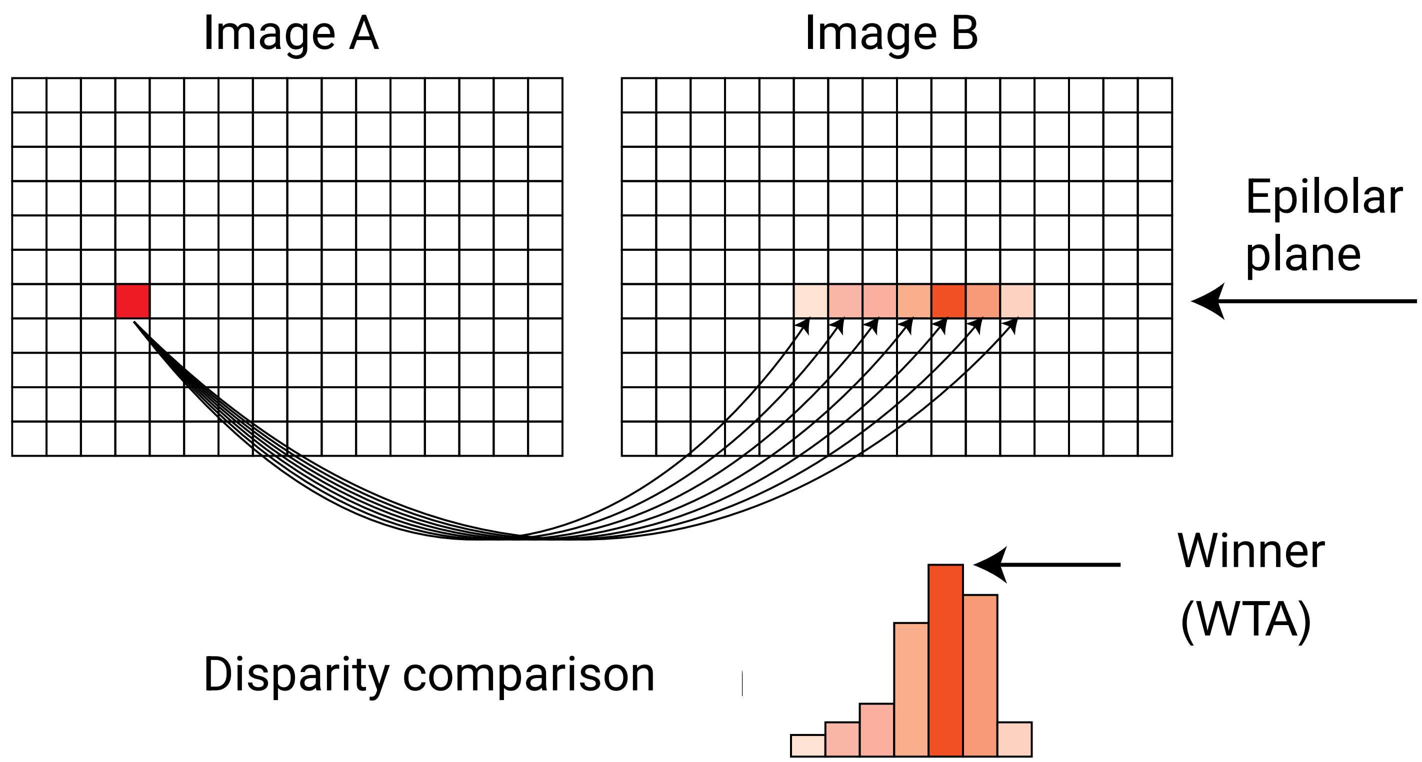


Depth image reconstruction

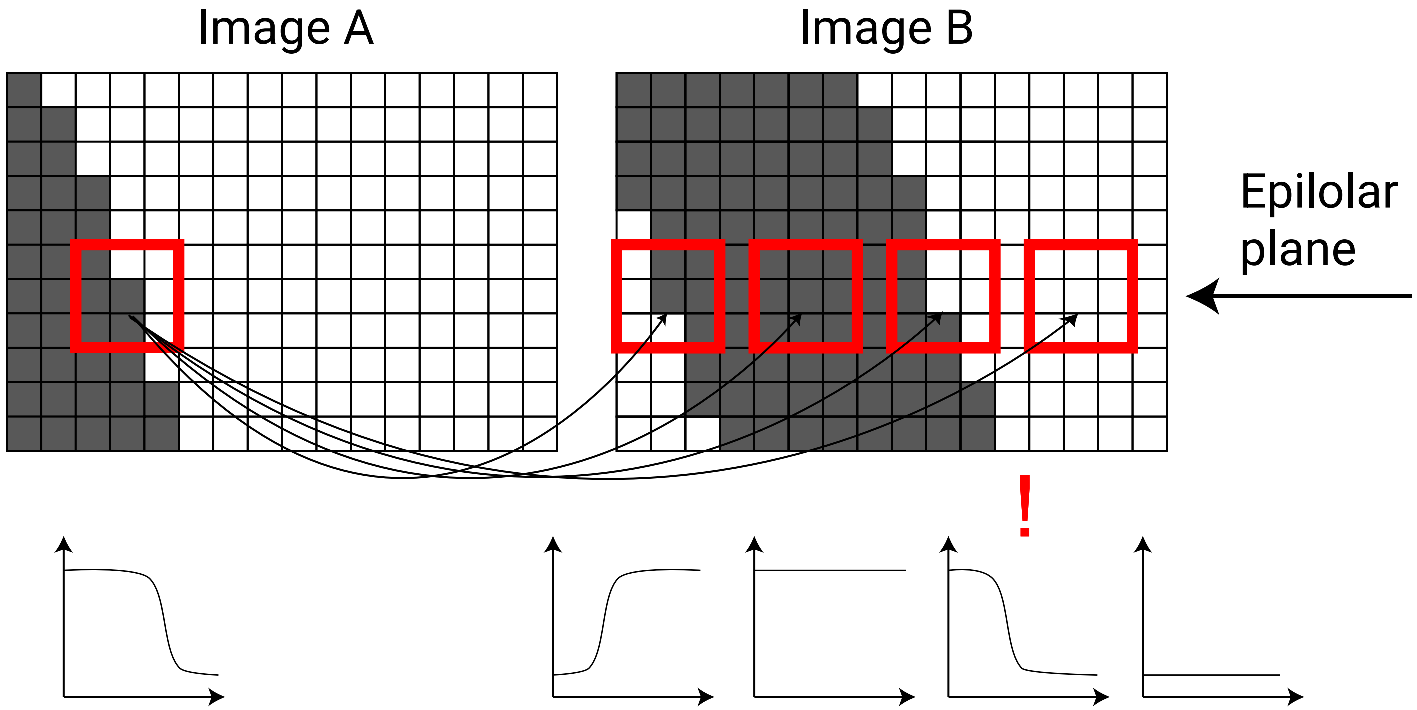


Correspondence Problem

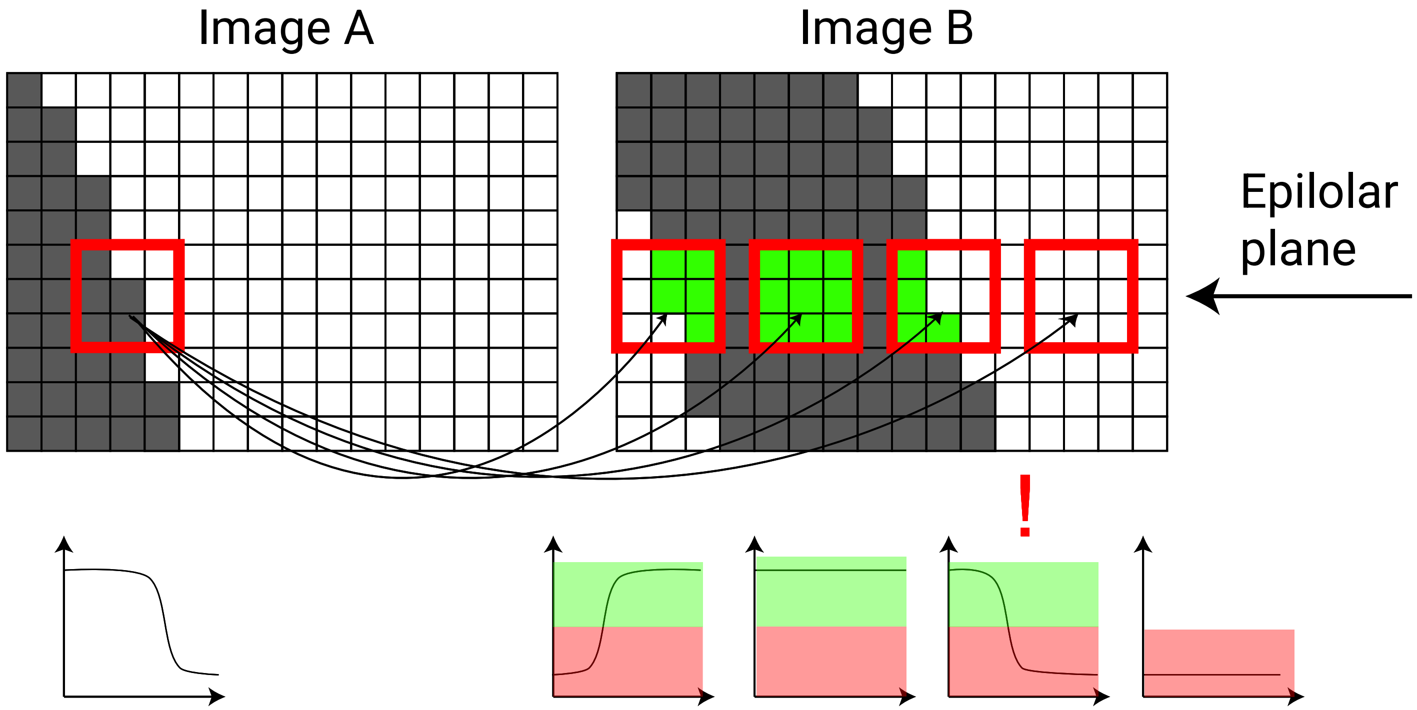
Naive Approach



Support Window



Support Window



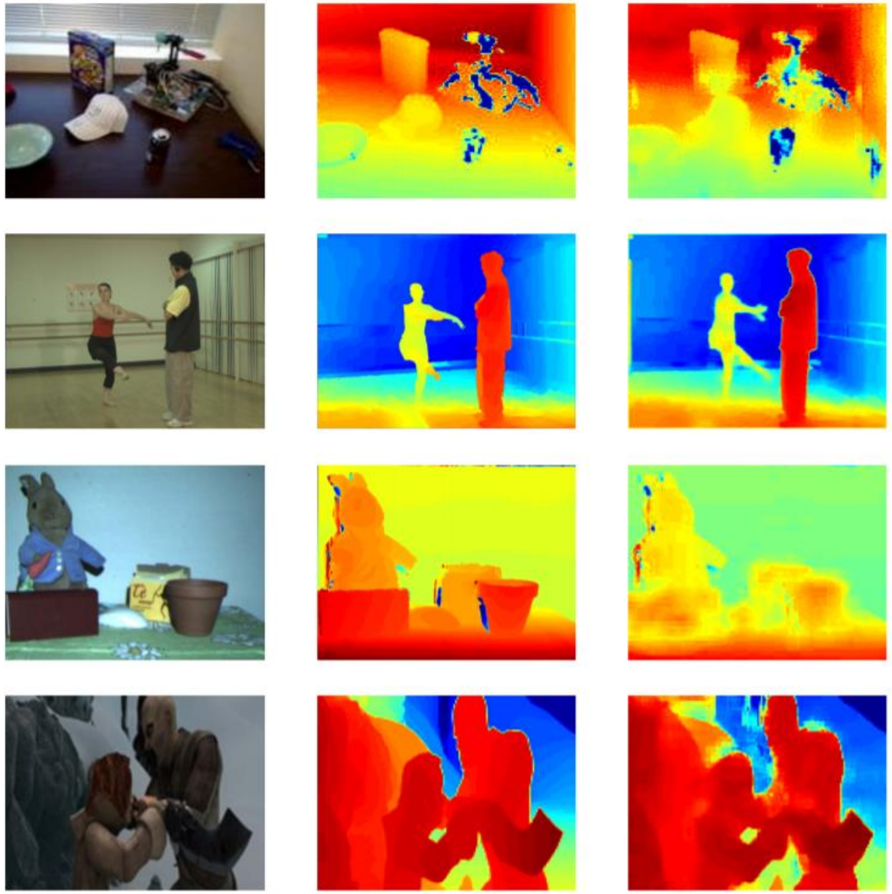

TEN TEN
Hubway
new balance

Correspondence

- **Local procedures**
  - Basically based on the "naïve" comparison (WTA)
  - Calculate for a limited area of the image
  - Smooth/filtered to avoid ambiguity
  - Examples: Block Matching, Support Window, Gradient-based Optimization, Local Feature Matching
- **Global (and semi-global) procedures**
  - Disparity comparisons across the entire image pairs
  - Minimize a cost function
  - Examples: Graph Cuts, Nonlinear Diffusion, Belief Propagation, Deep Reconstruction

Deep Reconstruction

- Reconstruction of depth images
- Neural networks are trained by single and depth image pairs
- Through transferred learning, the developed models can generate suitable depth images for new individual images in real time.

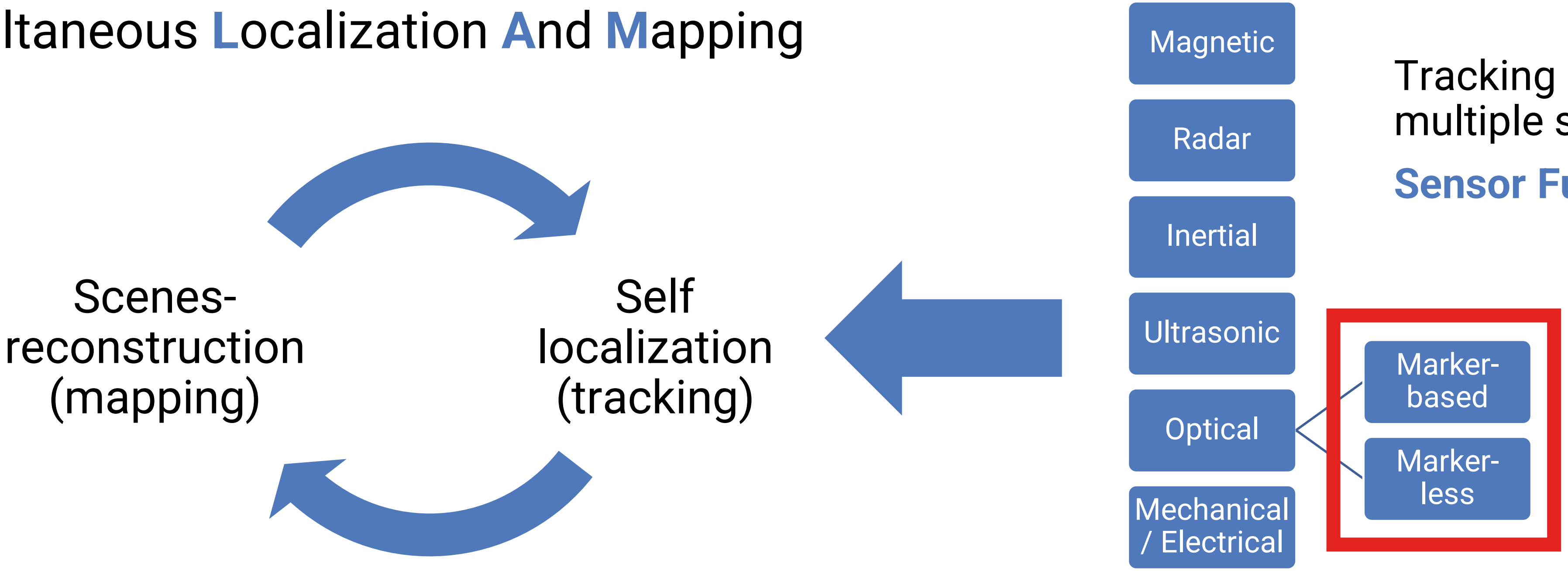


(a)Test Images (b)GT depth (c)FP depth

Nidhi Chahal, Meghna Pippal, and Santanu Chaudhury. 2016. Depth estimation from single image using machine learning techniques. In Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP '16). ACM, New York, NY, USA, , Article 19 , 8 pages. DOI: https://doi.org/10.1145/3009977.3010019

SLAM Problem

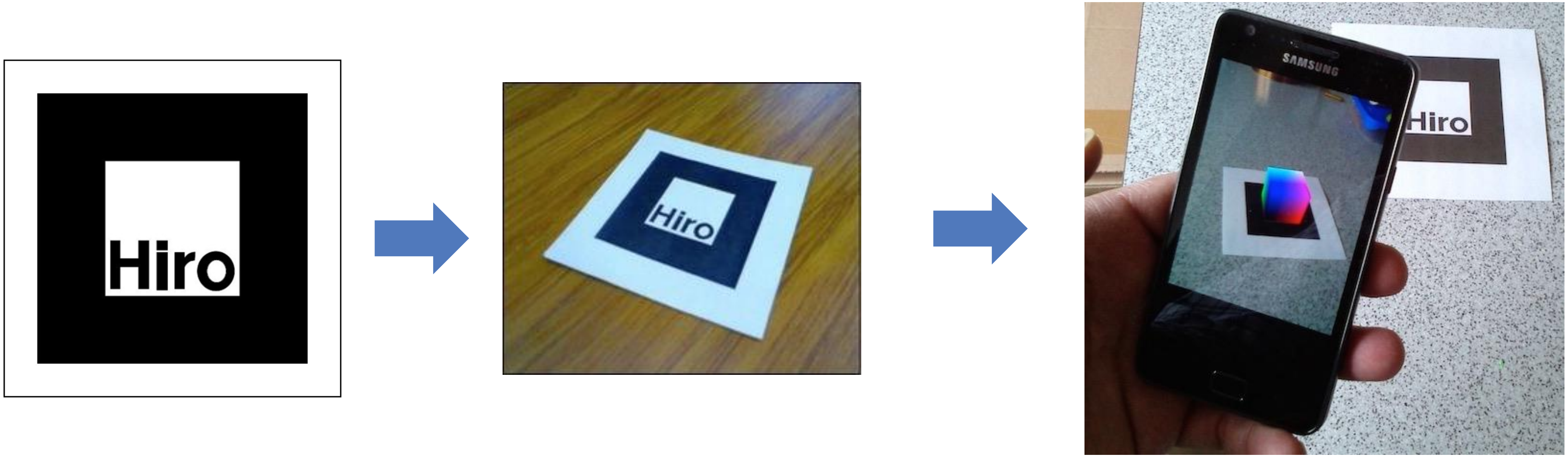
- Virtual and real objects must be properly matched, otherwise the illusion that they coexist will be broken
  - in particular, the movement of the device must be known
- **S**imultaneous **L**ocalization **A**nd **M**apping



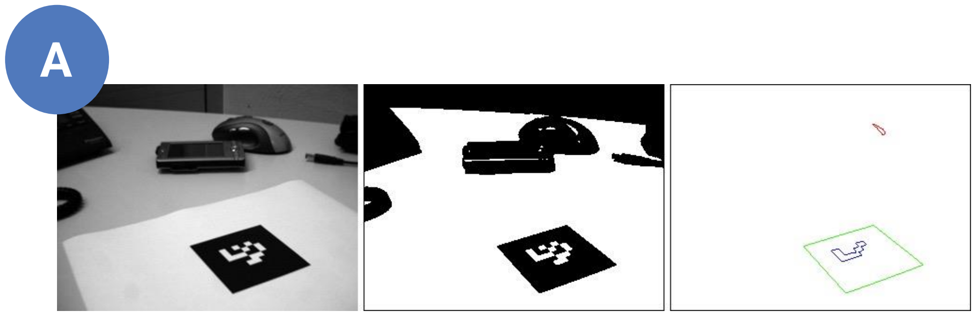
Tracking sensors for multiple sources:

Sensor Fusion

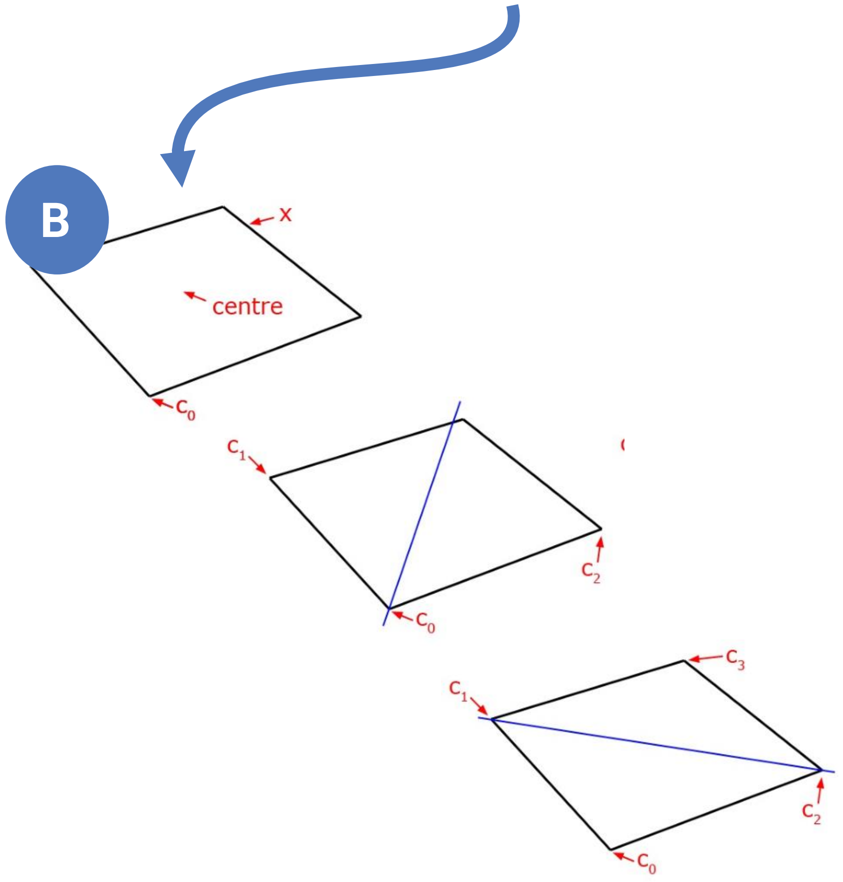
Marker-based Tracking



Marker-based Tracking

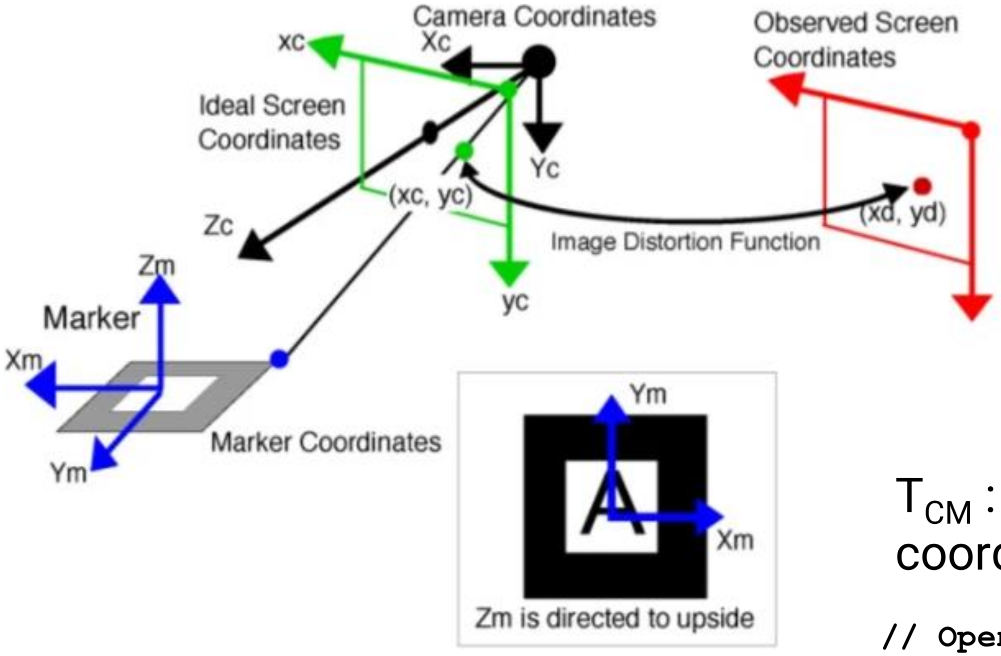


- **Step A: Fiducial Detection**
  1. set the picture to b/w
  2. search line by line for edges
  3. follow all edges either to the start pixel or to the edge of the image
  4. ignore edges that are too small (or too large)
- **Step B: Rectangle Fitting**
  1. start at any point x on one of the edges
  2. the point with the maximum distance to it is the corner c0
  3. create a diagonal through the middle
  4. find the points c1 & c2 with the maximum distance left and right to the diagonal
  5. determined the final diagonal from c1 to c2 and find point c3 to the right of the diagonal



From Marker to Camera

- Calculate the perspective transformation (rotation and translation) from the marker level to the camera level (homography)



$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_1 \\ R_{21} & R_{22} & R_{23} & T_2 \\ R_{31} & R_{32} & R_{33} & T_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix} = \mathbf{T}_{\mathbf{CM}} \begin{bmatrix} X_m \\ Y_m \\ Z_m \\ 1 \end{bmatrix}$$

T_{CM} : 4x4 Transformation matrix from marker to camera coordinates

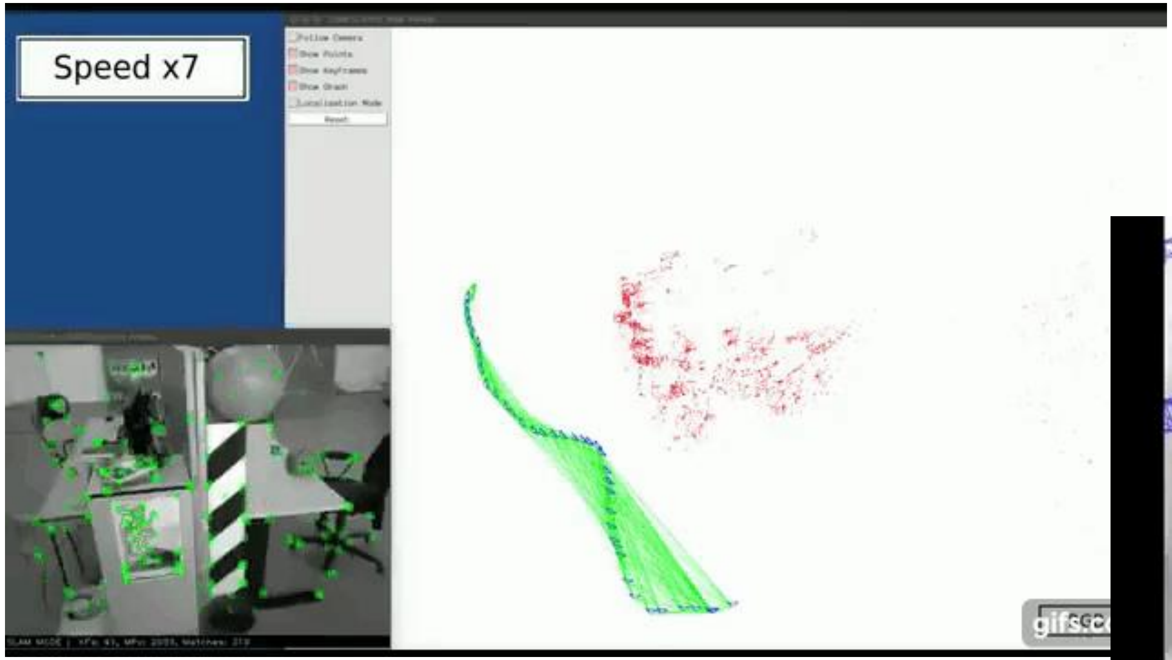
```
// OpenCV:
Mat cameraMatrix = findHomography(markerCoordinates, screenCoordinates);

// ARToolkit:
arGetTransMat(markerCoordinates, screenCoordinates, markerSize, cameraMatrix)
```

Marker-less Tracking

- **Basic SLAM approach**:
  - Track a series of points through successive camera images
  - Use the tracks to triangulate the 3D position of the points
  - Use the path of all point positions to calculate the camera pose
  - By observing a sufficient number of points, structures and movement (camera path and scene structure) can be reconstructed
- **Parallel tracking** and mapping by two concurrent processes (thread) for each frame

Marker-less Tracking



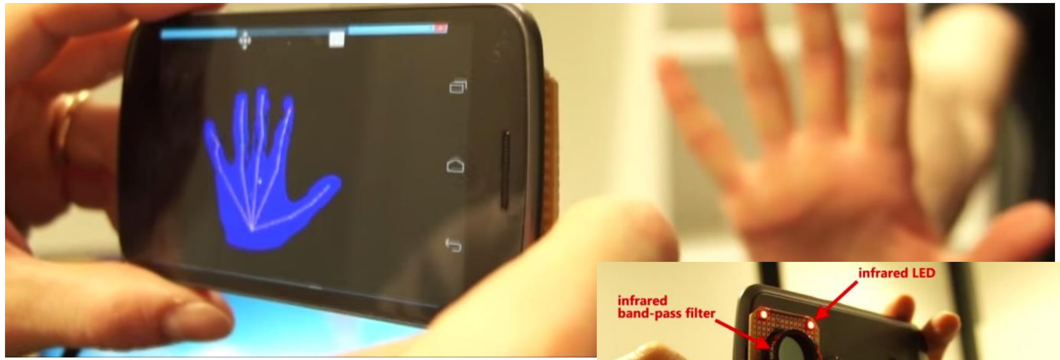
Mur-Artal, R., & Tardós, J. D. (2017). Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras. *IEEE Transactions on Robotics*, *33*(5), 1255-1262.

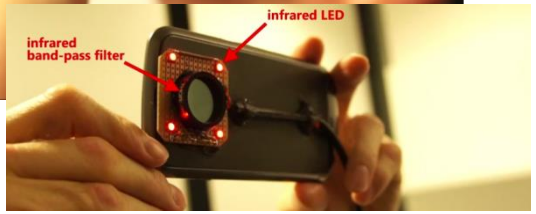
Sensor Fusion

- **Interial sensors (IMUs)**
  - often inaccurate
  - noisy output (= trembling augmentations) gives us first information about where we are in the world and what we are looking at
  - better temporal resolution (IMUs 16,000 Hz)
- **Cameras**
  - more accurate (= stable and correct augmentations)
  - requires the selection of the right key
  - point database from which to track
  - requires the registration of our local coordinate frame (online-generated model) to the global (world)temporally poorly resolved (30-90 Hz)

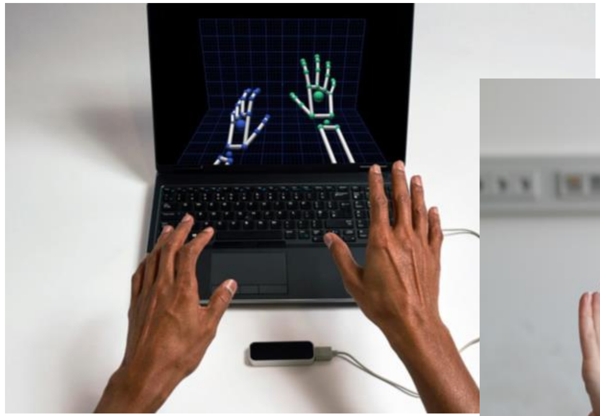
Let‘s combine them!

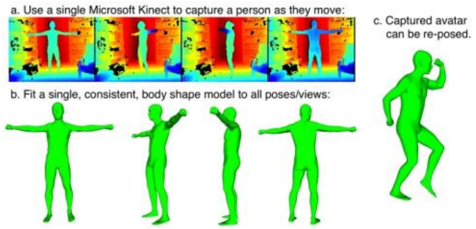
Hand and Body Tracking





Sean Ryan Fanello, Tim Paek, Cem Keskin, Shahram Izadi, Pushmeet Kohli, David Kim, David Sweeney, Antonio Criminisi, Jamie Shotton, and Sing Bing Kang. 2014. Learning to be a depth camera for close-range human capture and interaction. *ACM Transactions on Graphics* 33, 4: 1–11. https://doi.org/10.1145/2601097.2601223





Home 3D body scans from noisy image and range data. Weiss, A., Hirshberg, D. and Black, M.J. In *Int. Conf. on Computer Vision (ICCV)*, IEEE, pages 1951-1958, Barcelona. November 2011.

Method and Apparatus for Estimating Body Shape. Black, M.J., Balan, A., Weiss, A., Sigal, L., Loper, M. and St Clair, T. *US (12/541,898) and PCT patent application*, US (12/541,898) and PCT patent application, August 2009

Hand and Body Tracking

Oculus Rift

Magnus Glove + HTC Vive

OptiTrack + HTC Vive

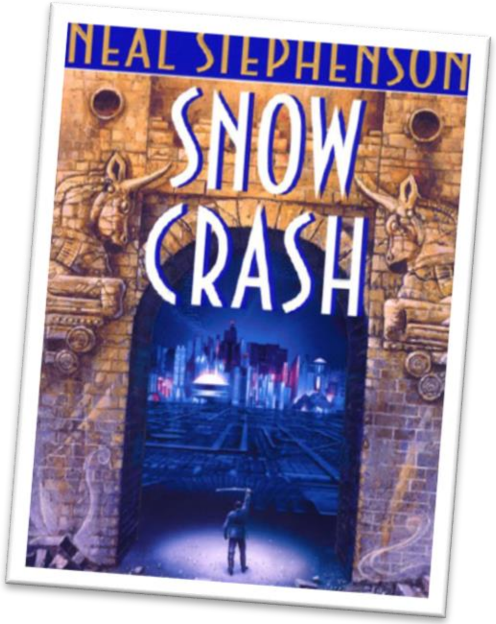

Alexei
Prayrit
Lucy

Interactions in Virtual Spaces

Human-Computer Interaction Lecture

Slides adapted from the Game Design Workshop by V. Schwind, Image from: https://pxhere.com/de/photo/764632

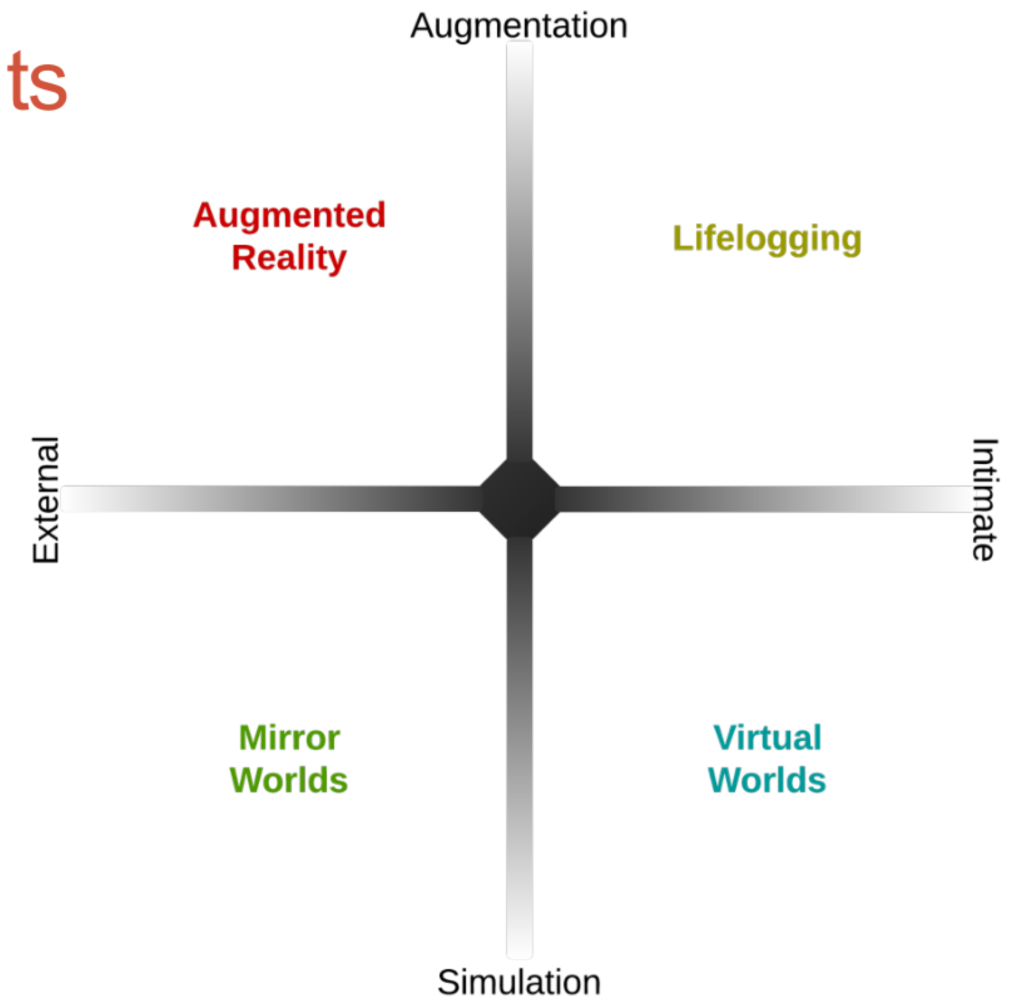
„Metaverse“



- Term from Neal Stephenson's "Snow Crash
- The metaverse is the convergence of
  - virtually enhanced physical reality
  - physically persistent virtual space
- Technical challenges in 4 dimensions:
  - Augmentation technologies that store information on our perception of the physical environment (scene registration)
  - Simulation technologies that model (and predict) reality
  - Intimate technologies that are inward and suggest the identity and actions of the individual or object
  - External technologies that are outward-looking and indicative of the world in general

Metaverse Components

- **Virtual Worlds** (Virtual Content)
- **Mixed Reality** (AR and VR)
- **Mirror Worlds**
  - Virtual replicas of the real world (Google Earth, Street View, Google Maps)
- **Lifelogging**
  - Capture and report the intimate states and life stories of objects and users with e.g., phyisiological sensors (FitBit, Apple watch,...)



Adapted from Smart, E.J., Cascio, J. and Paffendord, J., Metaverse Roadmap Overview

Metaverse: The Unknowns

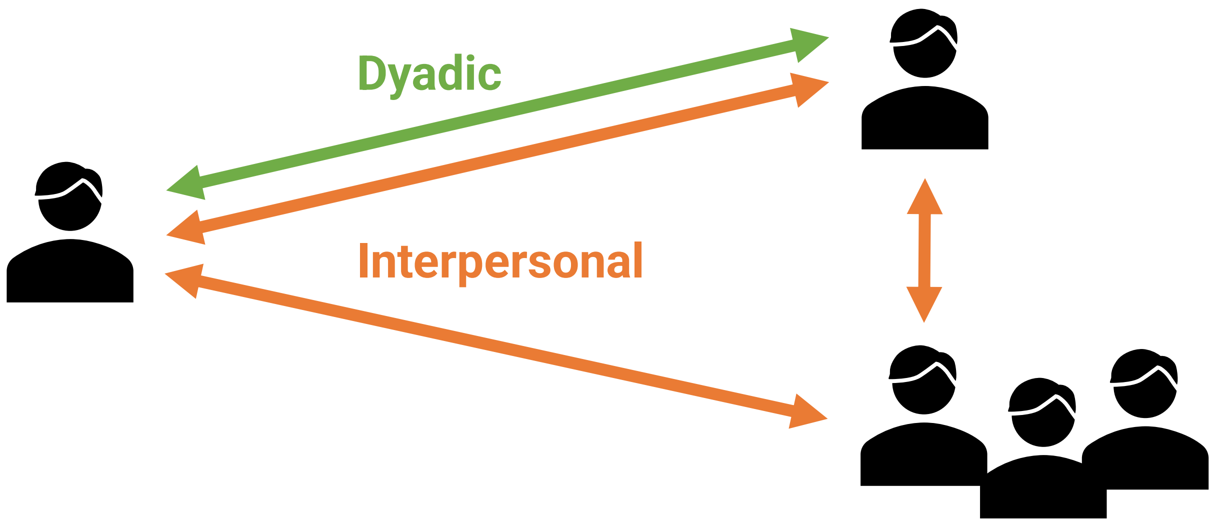
- **Legal and data protection issues**
  - What about hate, bullying, stalking, privacy?
  - What is (un)right in the metaverse?
- **Implications of Always-on Technologies**
  - How is what stored where?
  - Health issues?
  - Medical implications (diagonistics?)
- **Reliable models of the physical world**
  - Not always correct
  - Intimate states
- **Social perception**
  - …

Presence and Social Presence

- Recap: **presence** is the “sense of being and acting” in a simulated environment [1,2]
  - presence ≠ location
- **Social presence** or **Copresence** describe the “sense of being and acting with another person” in a simulated environment [3]
  - Prerequisite of social interaction and social influence induction
  - Users exhibit behavior in a simulated environment similar to a real-world
  - Enables **dyadic or interpersonal communication**

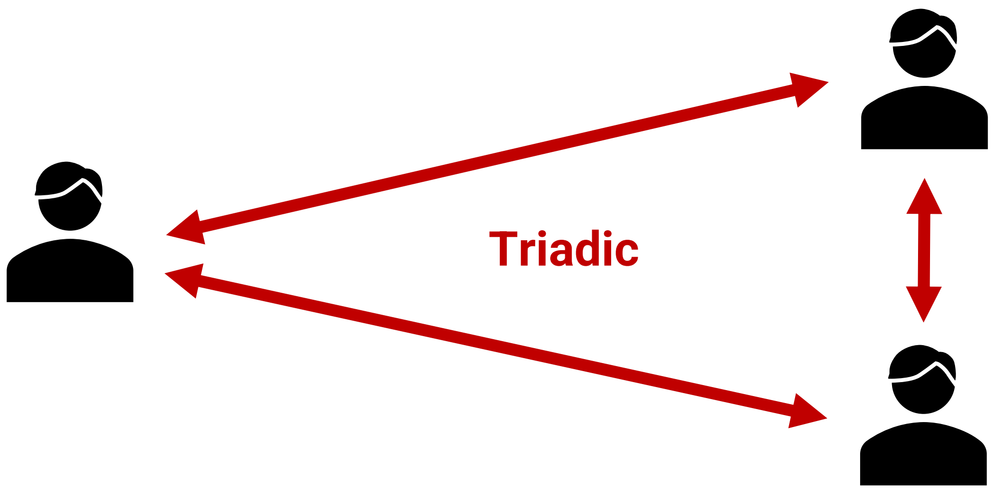
Dyadic and Interpersonal Interaction

- **Dyadic**: between two people
- **Interpersonal**: between two people or a larger group of people
- A sudden encounter between two strangers is not termed as interaction (no exchange of information) but can have social implications, of course.



Triadic Interaction

- **Triadic**: between three people
- A huge number of social relationships are not possible between two people (dyad), but only occur with three (triad).
  - coalition forming, intrigues, interpersonal secrets, “the impartial”, “the laughing third”, …

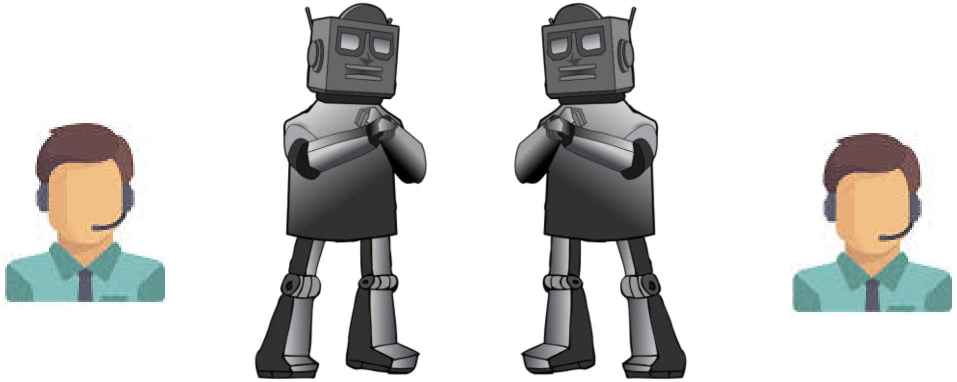


Presence Definition

		The other is located…	
		in **Physical Proximity**	in **Digital Proximity**
The other is present…	in **Person**	**Corporeal Copresence** (face-to-face) *e.g., Dining with someone at a restaurant*	**Corporeal Telecopresence** (face-to-device) *e.g., Talking to a friend over the phone*
	via **Simulation**	**Virtual Copresence** (instrumental or communicative robots) *e.g., Interacting with an ATM machine*	**Virtual Telecopresence** (instrumental or communicative virtual agents) *e.g., Talking with Alexa*

Hypervirtual (tele-)copresence

- When **persons** are present through simulation
  - **physically** in each other's physical proximity, we have **hypervirtual copresence**
  - **digitally** in each other’s digital proximity, we have **hypervirtual telecopresence**



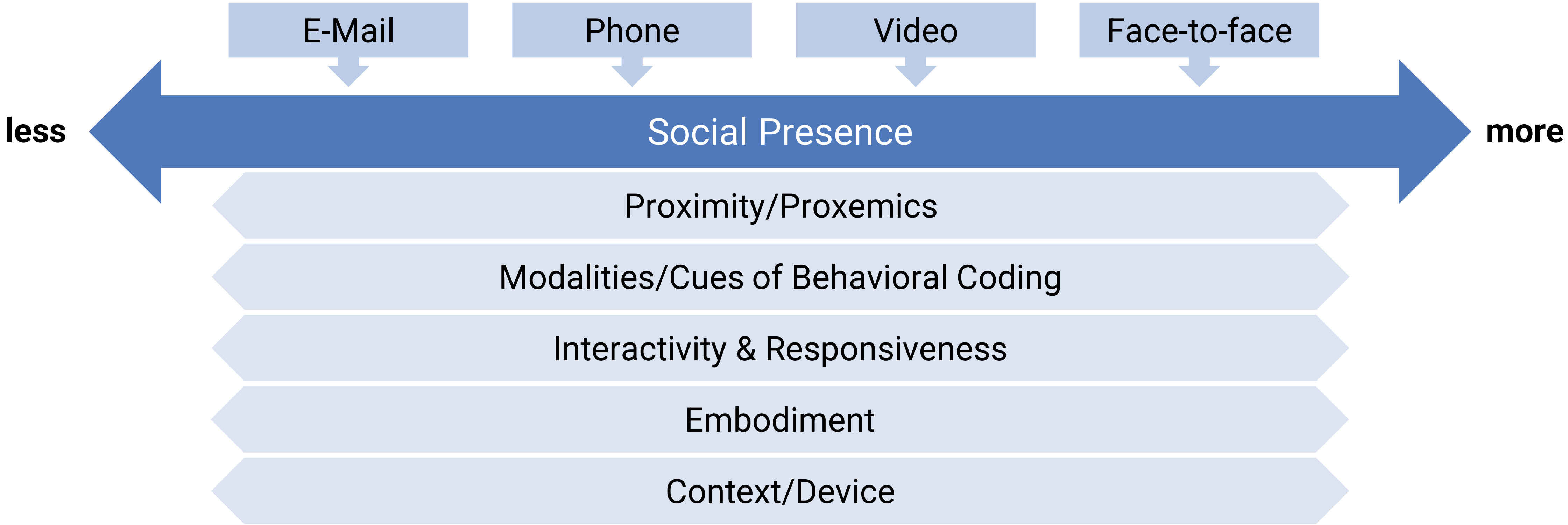
hypervirtual copresence

hypervirtual telecopresence

Influences on Social Interaction

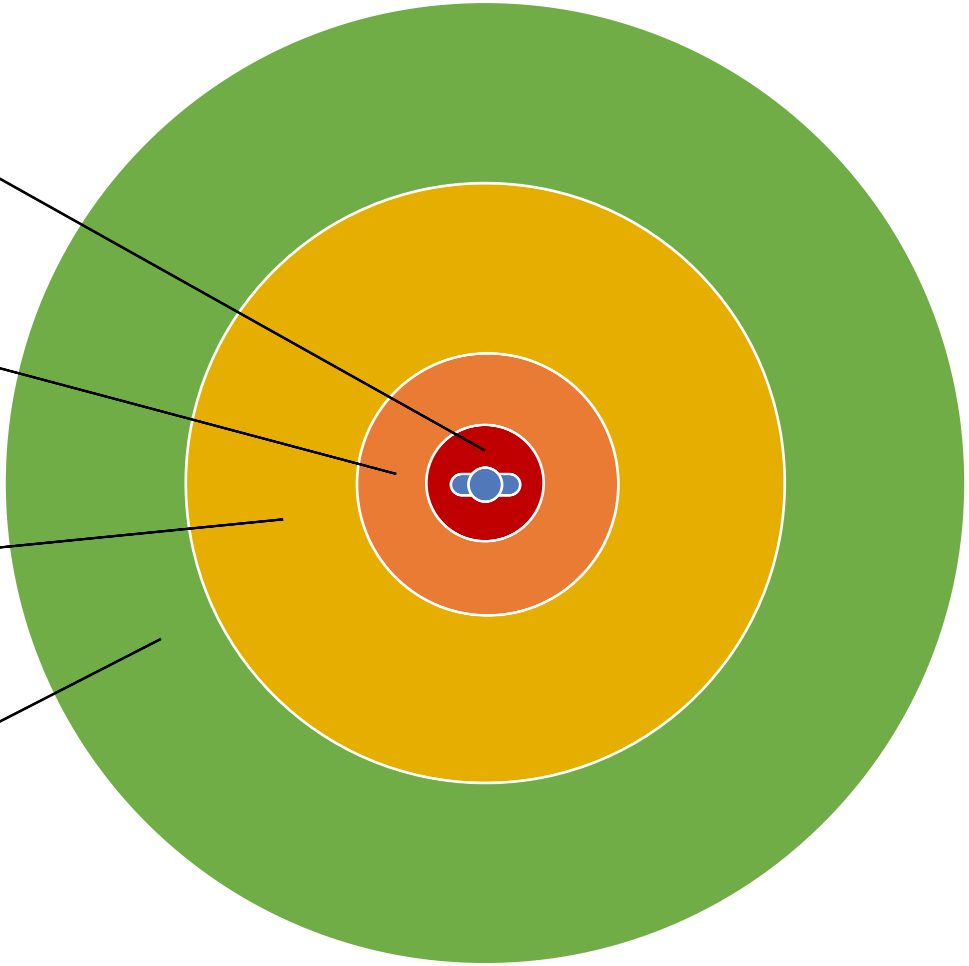
- The mere **digital** or **physical copresence** of others is sufficient **to induce social influence**
  - even the absence of people can have social implications
- Important parameters that influence social interaction
  - **Embodiment**: the involvement of human or human-like bodies
  - **Interactivity & Responsiveness**: the communication speed between copresent individuals
  - **Scale**: the number of copresent individuals interacting with each other
  - **Mobility**: the ability of the copresent individuals to carry the interaction
    - e.g., for devices in telecopresence : stationary (desktop computers), portable (cars, tablets), and wearable (head-mounted displays)

Multimodal Cues



Proxemics of Human Territory

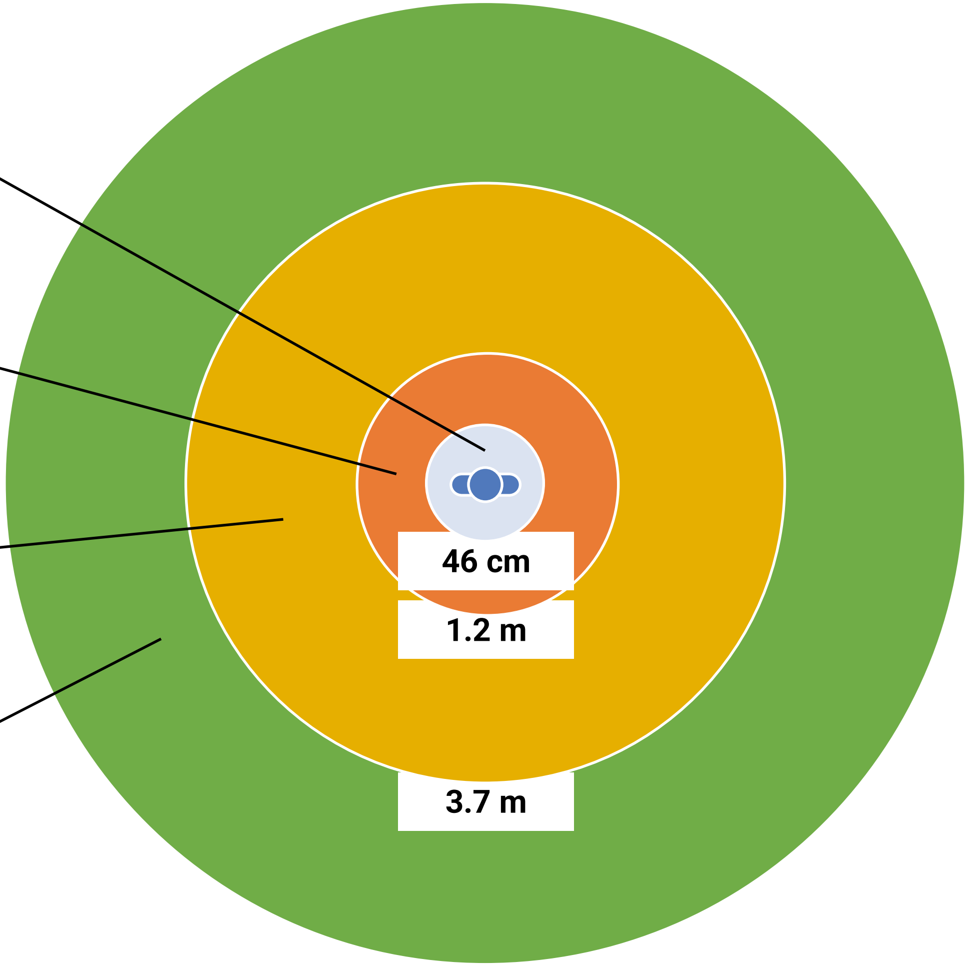
- **Body territory**
  - the local space immediately surrounding us (high emotional reactions)
- **Home territory**
  - the space where people continuously have control over their surrounding
- **Interactional territory**
  - a place where people congregate informally
- **Public territory**
  - space out of constant control, free entering



Hall, Edward T. (1966). The Hidden Dimension. Anchor Books. ISBN 978-0-385-08476-5.

Proxemics of Human Territory

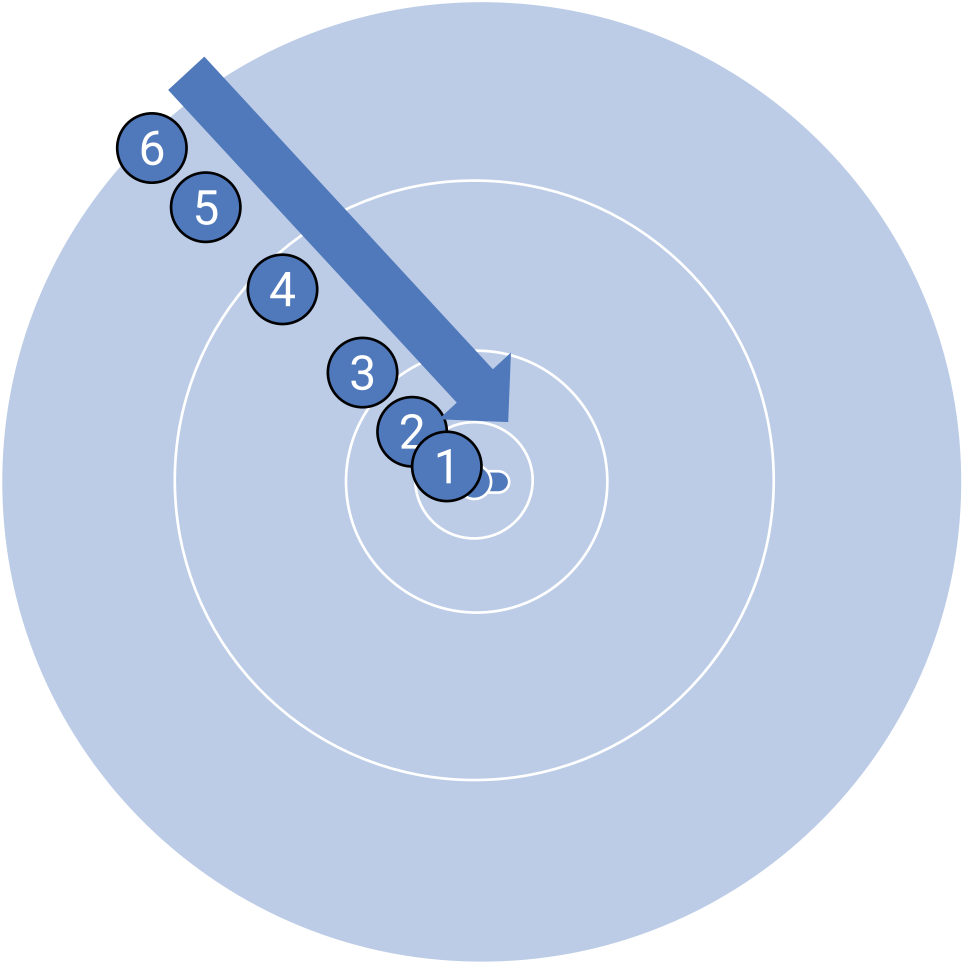
- **Intimate space**: 0 to 45 cm
  - space for embracing, touching or whispering among partners, close relatives
- **Personal space**: 46 cm to 1.2 m
  - space for interactions among good friends, siblings, other relatives
- **Social space**: 1.2 m to 3.7 m
  - space for interactions among acquaintances
- **Public space**: 3.7 m and higher
  - space for public speaking, talks, …



Hall, Edward T. (1966). The Hidden Dimension. Anchor Books. ISBN 978-0-385-08476-5.

Modalities/Cues of Behavioral Coding

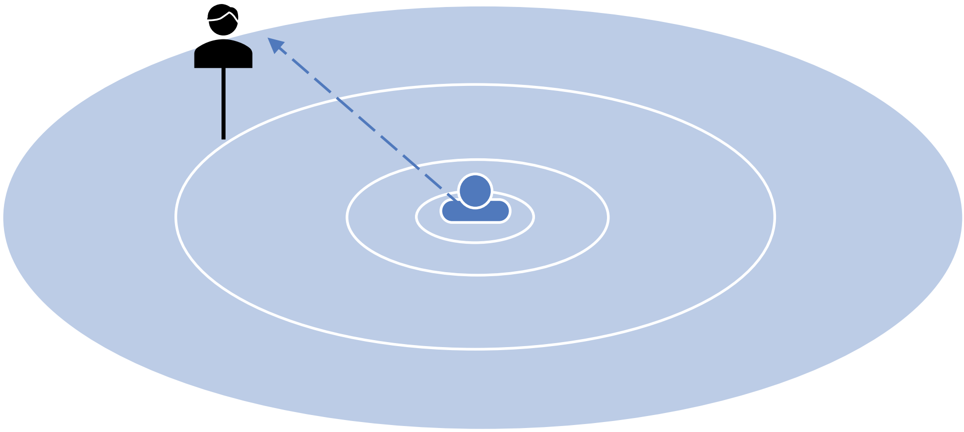
1. **Tactile & Haptic Codes**
   - how individuals are touching each other
2. **Thermal Codes**
   - the amount of body heat
3. **Olfactory Codes**
   - kind and degree of odor
4. **Kinesthetic Codes**
   - kind and degree of body language and facial expreesions
5. **Voice & Sound Codes**
   - speech and vocal effort
6. **Visual Codes**
   - external appearance, eye contact



Hall, Edward T. (1966). The Hidden Dimension. Anchor Books. ISBN 978-0-385-08476-5.

The Vertical Space

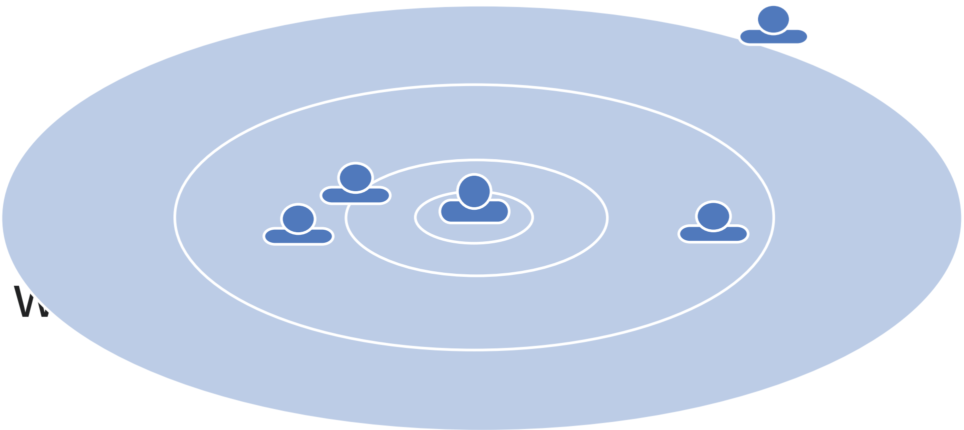
- **Height differences** between individual communicate the **degree of dominance** (or sub-ordinance) in a social relationship
  - e.g., adults vs children, king vs folks
- **Higher height** indicate **higher status** and more control over the relationship and the territory
- **Lower height** indicate **lower status** and less control over the relationship and the territory



Hall, Edward T. (1966). The Hidden Dimension. Anchor Books. ISBN 978-0-385-08476-5.

Social Isolation & Loneliness

- **Repeated exposure** to any stimulus **can decrease the mental effort** with which it is cognitively being processed [1]
  - **reduced mental effort increases positive affect**
  - requires positive enforcement (repeated observations)
  - works with humans, objects, animals, food,…
- Does not necessarily require (but works much better w
  presence [2]
- Getting used to social contacts can lead to **negative emotions if they are absent**



[1] Moreland & Zajonc (1982): Exposure effects in person perception: Familiarity, similarity, and attraction. Journal of Experimental Social Psychology, 18, S. 395–415
[2] O'Leary, Michael Boyer; Wilson, Jeanne M.; Metiu, Anca; Jett, Quintus R. (2008). "Wahrgenommene Nähe in der virtuellen Arbeit: Erklärung des Paradoxons von weit, aber nah" . Organisationsstudien . 29 (7): 979–1002. doi : 10.1177/0170840607083105 . S2CID 7715386 .

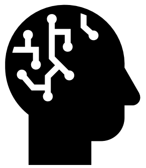
How Do People Meet?

- **Collaboration & Cooperation**
  - Sharing knowledge: teaching, training, learning, instructing, …
  - Sharing properties: trading, exchanging, storing …
  - Solving problems: working, servicing, searching, …
- **Strengthening of Social Bonds**
  - Family, friendship, colleagues, social pressure …
- **Games & Sports**
  - Competition, fun, self-monitoring, …
- **Confrontation**
  - Harming, bullying, defamation, annihiliaton (don‘t do this) …
- **By Accident**

How Do People Communicate?

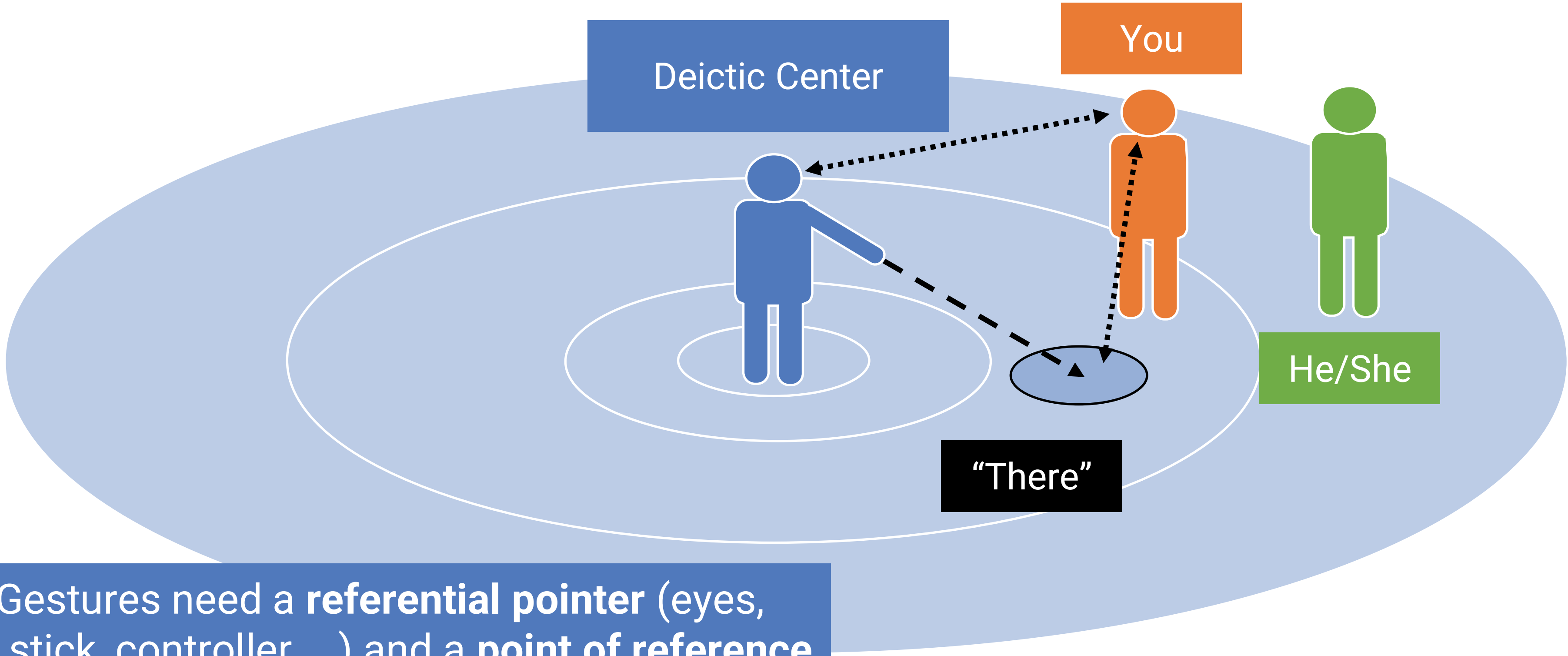
- **Verbal/Auditive Cues**
  - Speech
  - Vocalics
  - Phonemics
  - …
- **Nonverbal Cues**
  - Textual, Symbolic
  - Body Language
  - Mimics, Gestures
  - …

coded → **Signs** → interpreted → **Meaning**

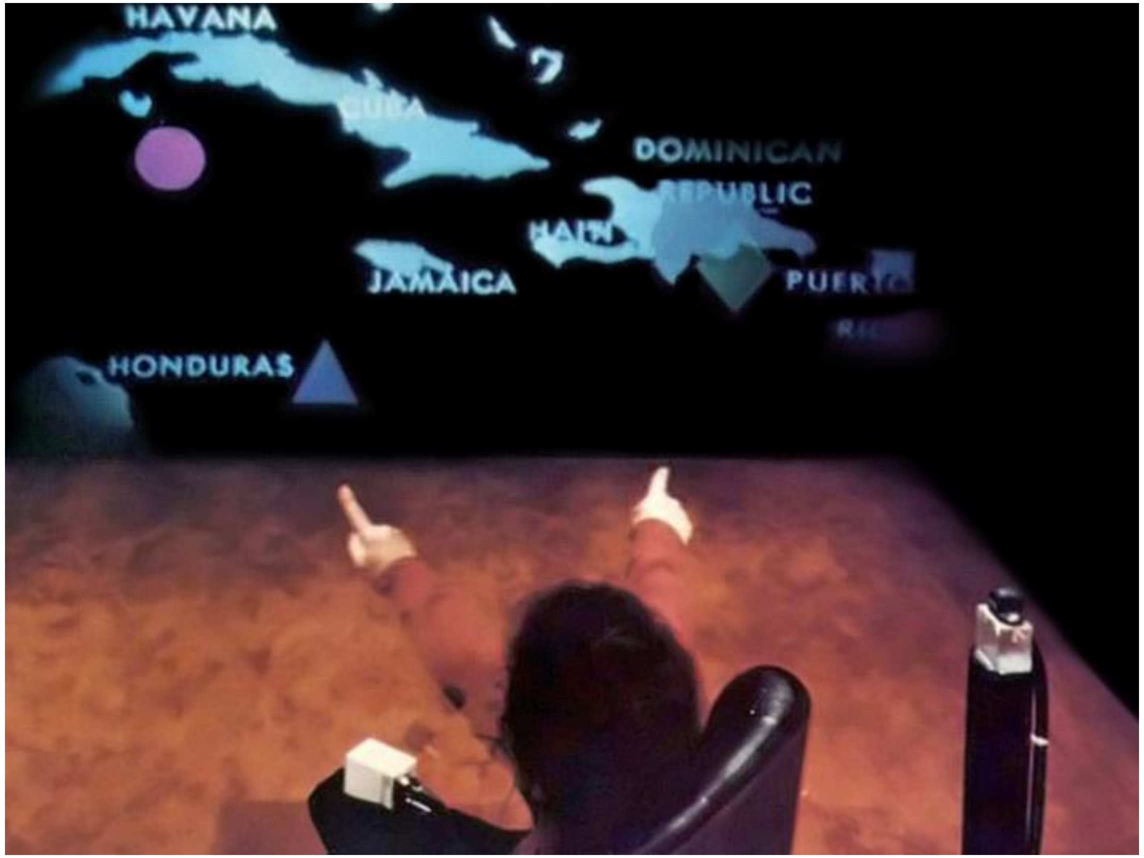


Deixis

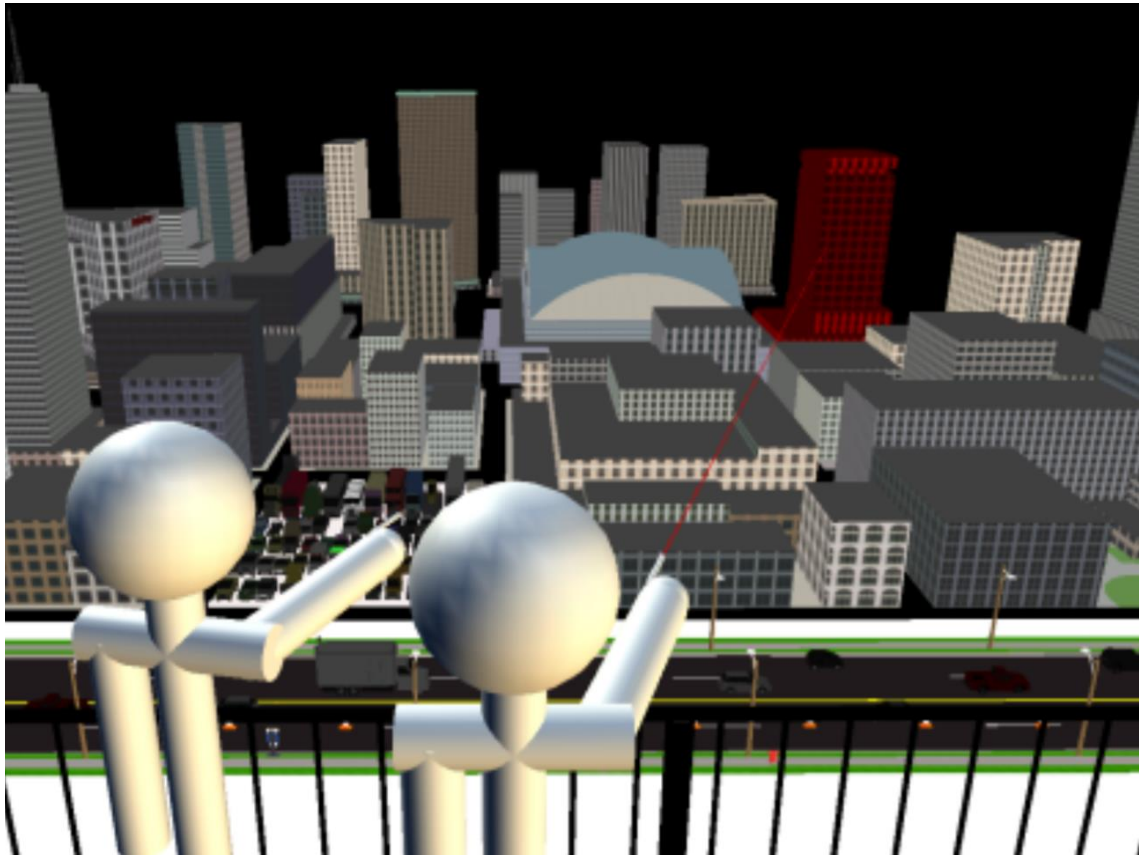
Deixis in Social Virtual Spaces



→ Deictic Gestures need a **referential pointer** (eyes, hand, arm, stick, controller, …) and a **point of reference** (object, human, …)

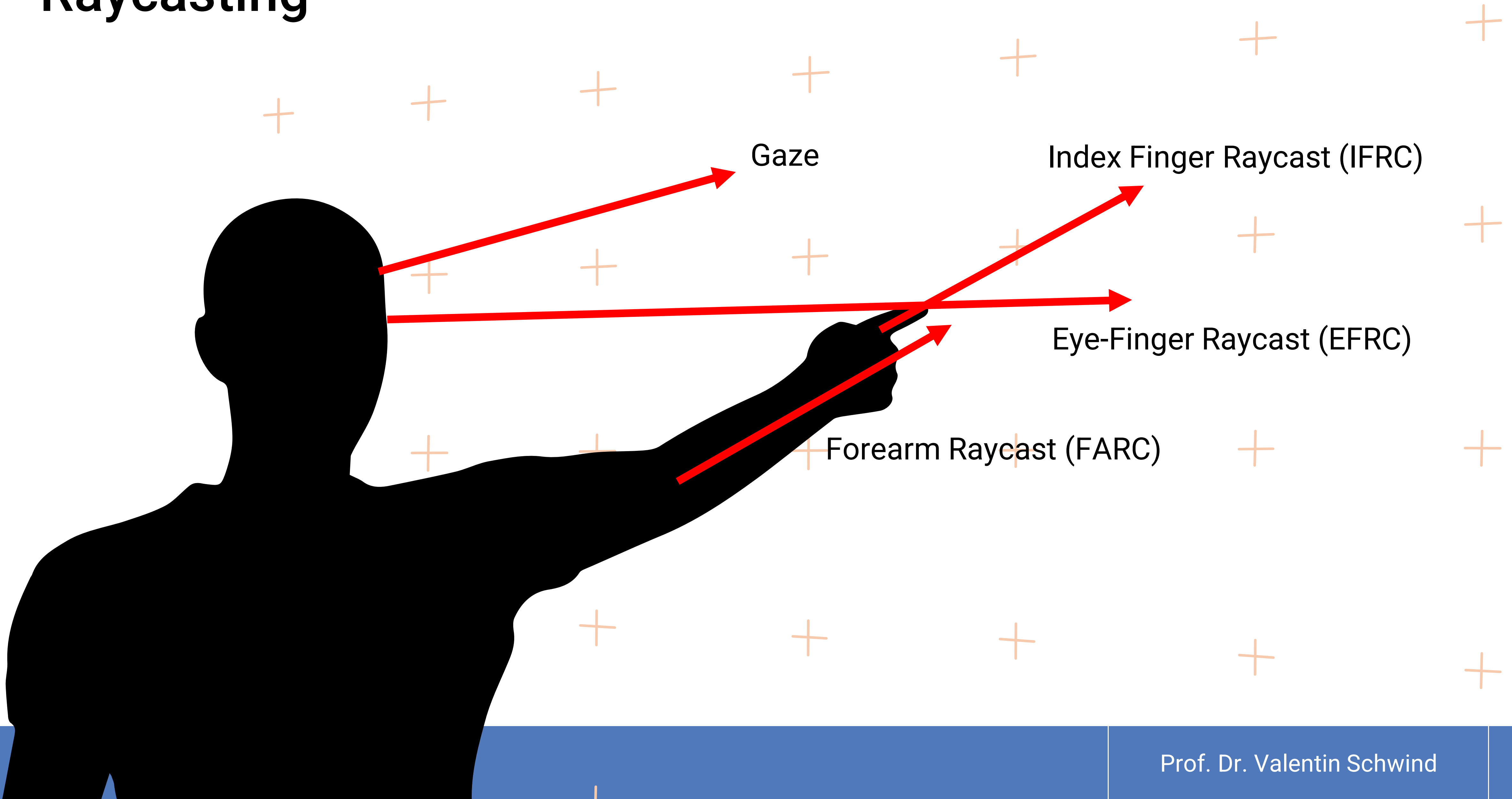


Richard A. Bolt. 1980. “Put-that-there”: Voice and gesture at the graphics interface. In Proc. of SIGGRAPH '80. DOI: http://dx.doi.org/10.1145/800250.807503

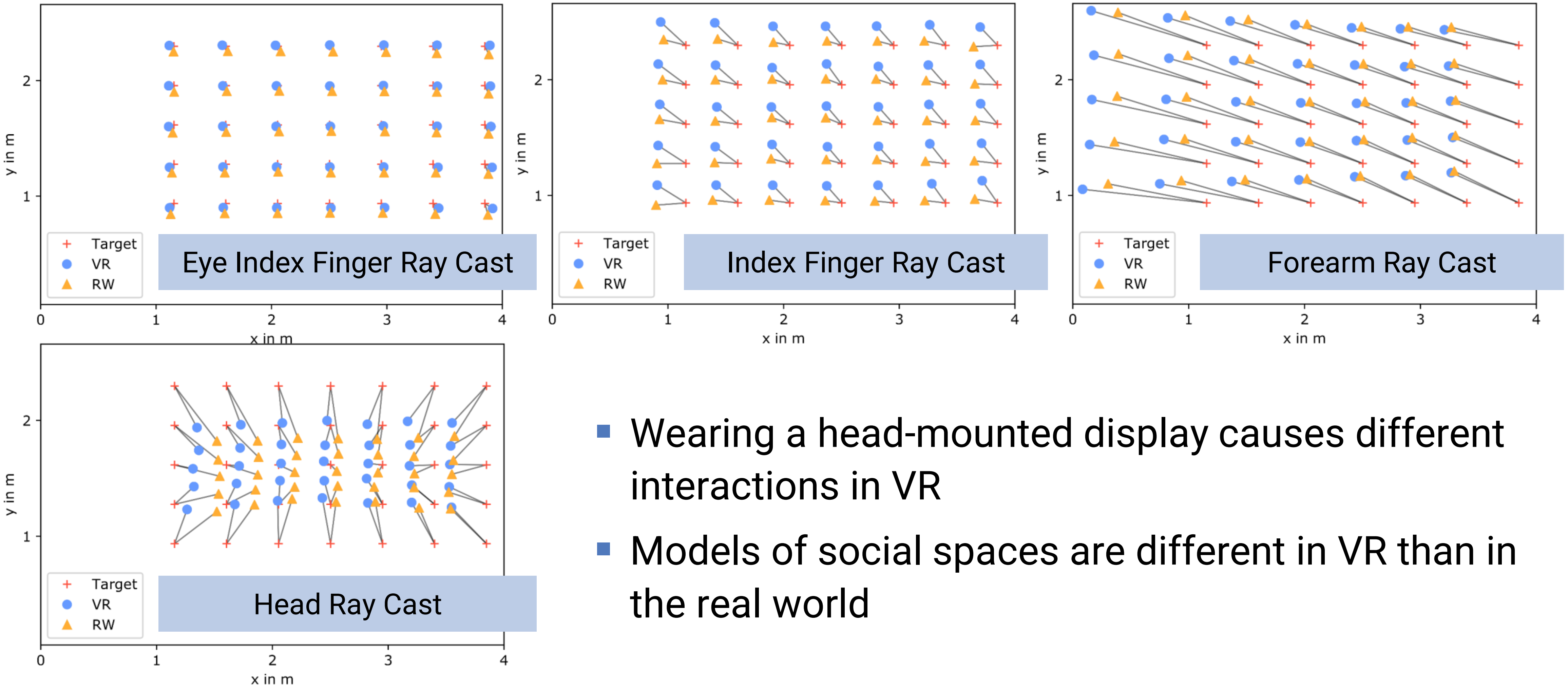


Nelson Wong and Carl Gutwin. 2014. Support for deictic pointing in CVEs: still fragmented after all these years'. In Proc. of CSCW '14. DOI: https://doi.org/10.1145/2531602.2531691

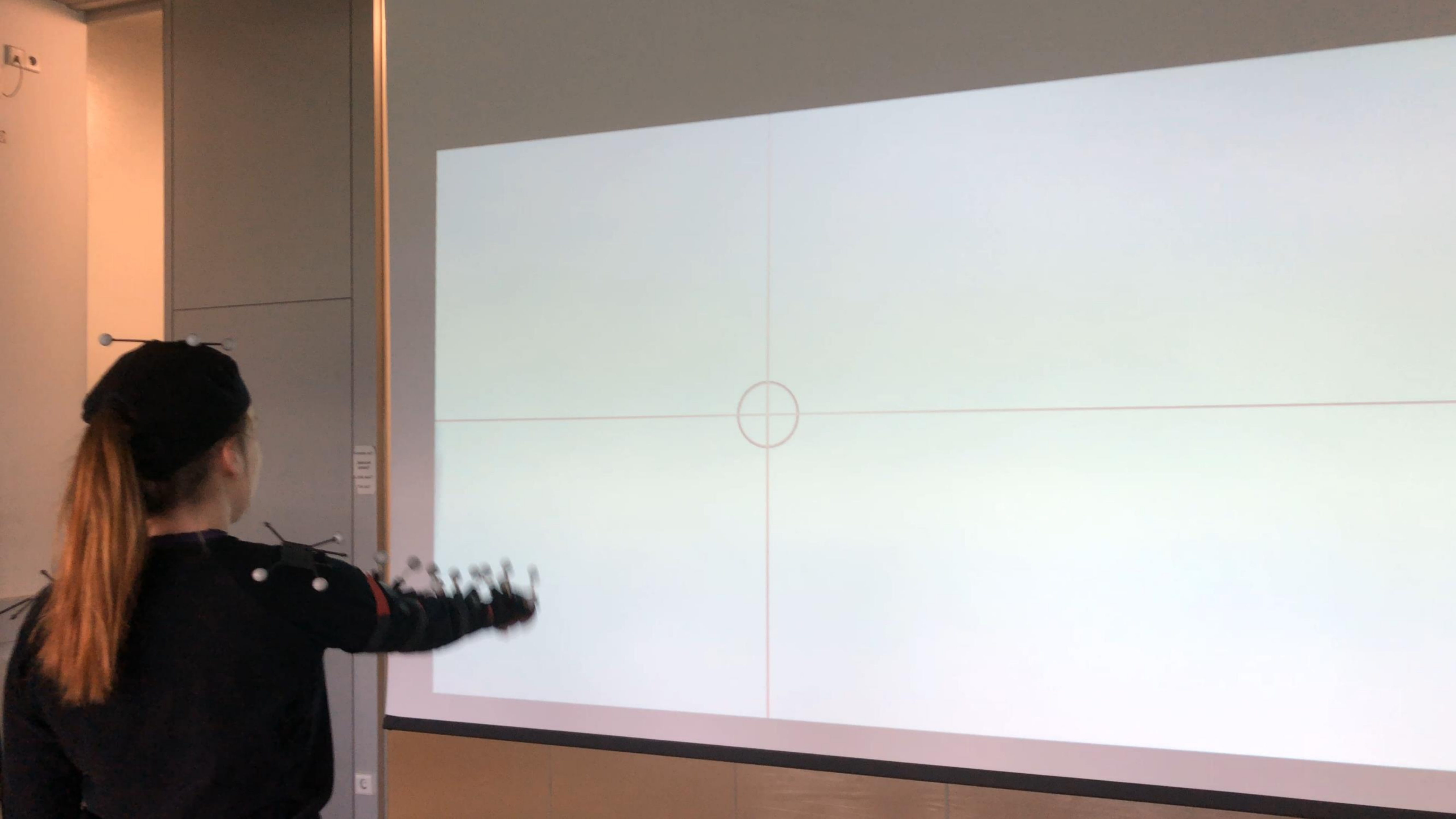
Raycasting


Gaze
Index Finger Raycast (IFRC)
Eye-Finger Raycast (EFRC)
Forearm Raycast (FARC)

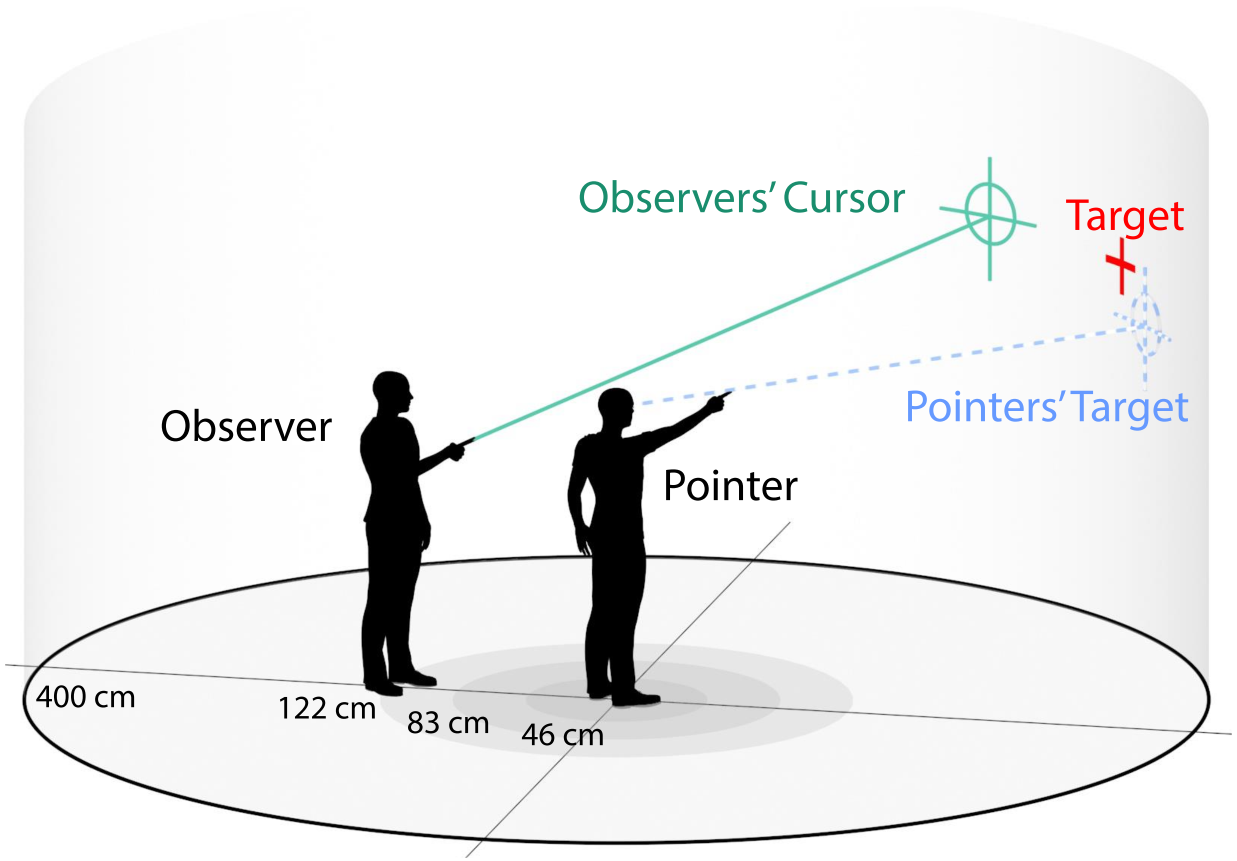
Raycast Differences in Virtual Reality

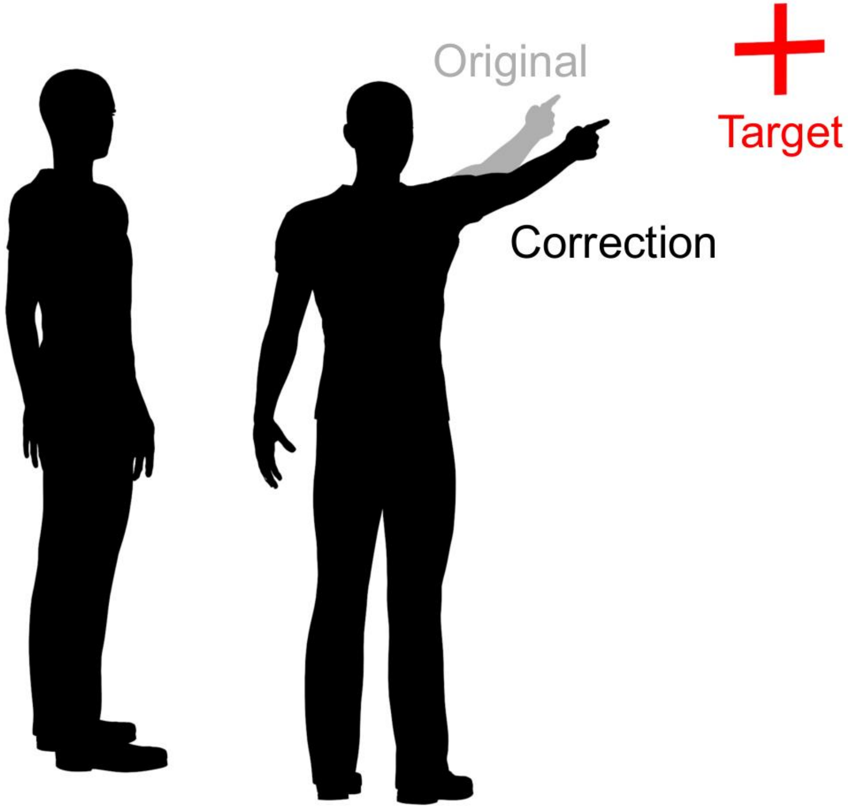


- Wearing a head-mounted display causes different interactions in VR
- Models of social spaces are different in VR than in the real world



Deictic Proxemic Behavior



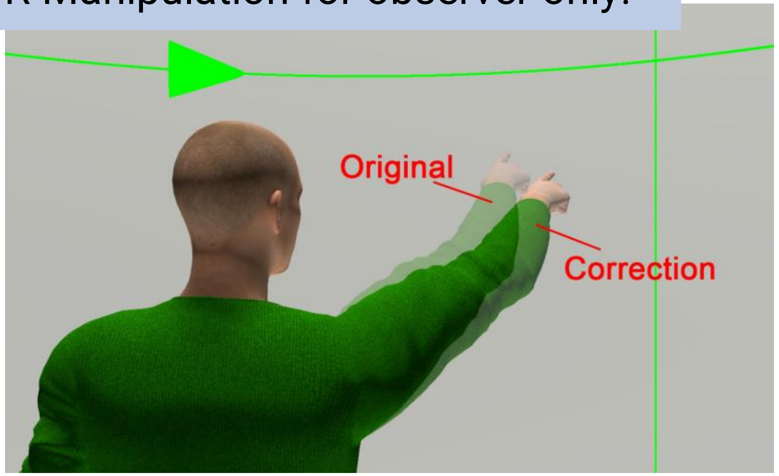


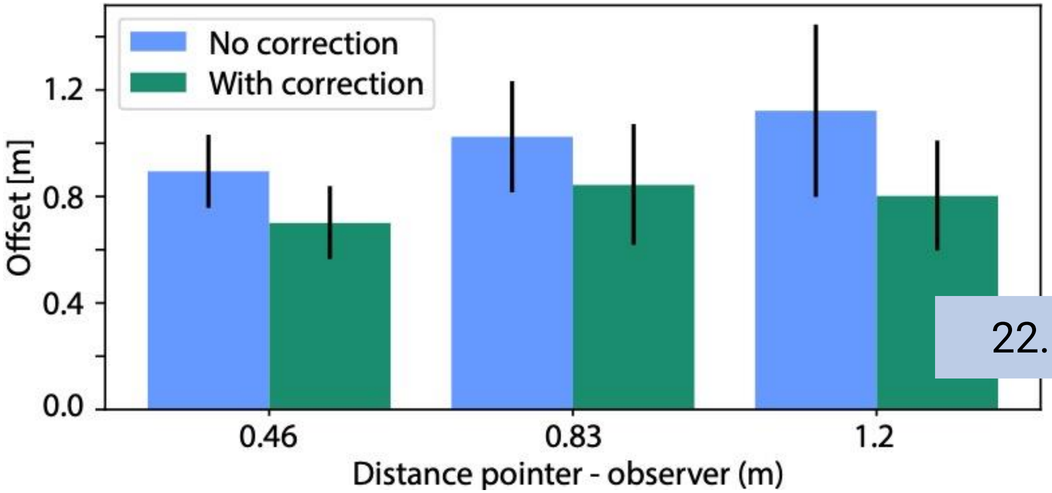
Sven Mayer, Jens Reinhardt, Robin Schweigert, Brighten Jelke, Valentin Schwind, Katrin Wolf, and Niels Henze. 2020. Improving Humans' Ability to Interpret Deictic Gestures in Virtual Reality. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376340

Correction Model

$$\Delta_\omega = f_\omega(\alpha_p, \alpha_y) = (sin(freq_p(\alpha_p - pshift_p))amp_p + offset_p + 1) * (sin(freq_y(\alpha_y - pshift_y))amp_y + offset_y + 1)$$

VR Manipulation for observer only!



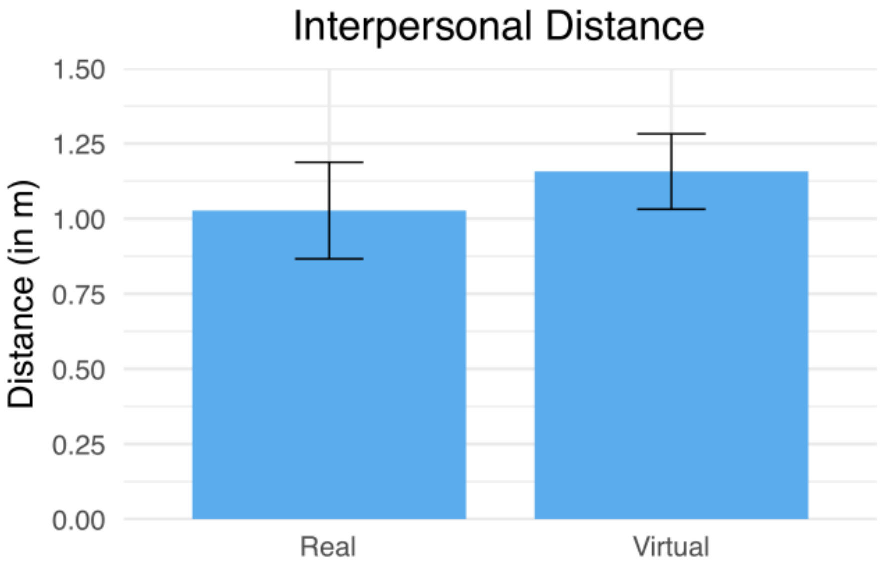


22.9% more accuracy

Sven Mayer, Jens Reinhardt, Robin Schweigert, Brighten Jelke, Valentin Schwind, Katrin Wolf, and Niels Henze. 2020. Improving Humans' Ability to Interpret Deictic Gestures in Virtual Reality. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376340

Interpersonal Distance

- Humans increase their interpersonal distance in VR
  - Halls Proxemics are not longer valid in VR
- Reason yet unknown (fear of colliding, odd-looking avatars,…)



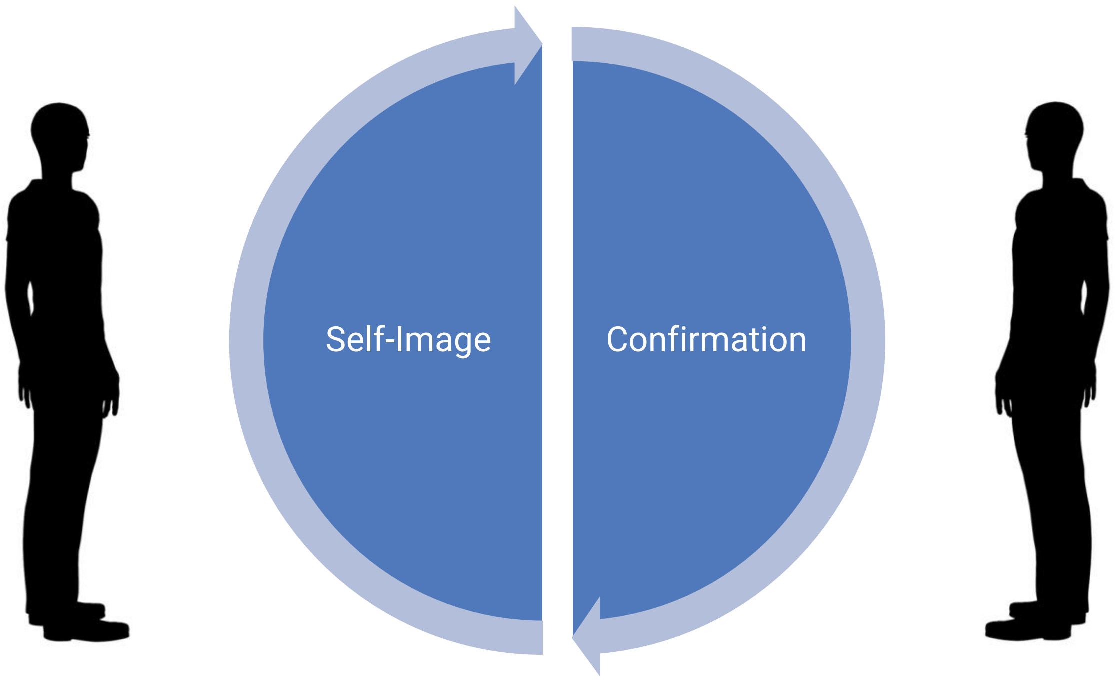
R. Rzayev, F. Habler, P. Ugnivenko, N. Henze, and V. Schwind. 2017. “It’s Not Always Better When We’re Together: Effects of Being Accompanied in Virtual Reality” Submitted to CHI ‘20

Avatars

- **Avatars in virtual reality** can manipulate the way how we perform **communication**
  - Observations known from the real world cannot be transferred 1:1
  - Proxemics must be refined (yet unknown)
- **However, the manipulation in VR** can be used **to improve non-verbal communication** between multiple people
  - Not possible in the real world
- Do different avatars change the way how we perform interactions?

Self-Image using Avatars

- **People adapt their behavior** according to their avatar (Proteus Effect) [1]
- **The more attractive** avatar, **the more confident** the user‘s behavior [2]
- Social implications of having another appearance?
  - Unexplored
  - Not possible in real



Self-Image using Avatars

- **The presence of a human** stereotype can **stimulate** the cognitive or physiological **performance**
- **Social self-image confirmation theory**:
  - Social collaboration highly depends on the extent to which partners agree with each other's self-image