IEEE VR 2018 REUTLINGEN

Christian Richardt

Motion-Aware Displays

IEEE VR Tutorial on Cutting-Edge VR/AR Display Technologies

CAMERA
Centre for the Analysis of Motion, Entertainment Research and Applications

UNIVERSITY OF BATH

Schedule

Start	Topic	Speaker
09:00	Introduction	George Alex Koulieris
09:30	Multi-focal displays	George Alex Koulieris
10:30	Coffee break	
11:00	Near-eye varifocal AR	Kaan Akşit
12:00	Lunch	
14:00	HDR-enabled displays	Rafał Mantiuk
14:45	Gaze-aware displays	Katerina Mania
15:30	Coffee break	
16:00	Motion-aware displays	Christian Richardt
17:00	Panel	All presenters

Why care about motion?

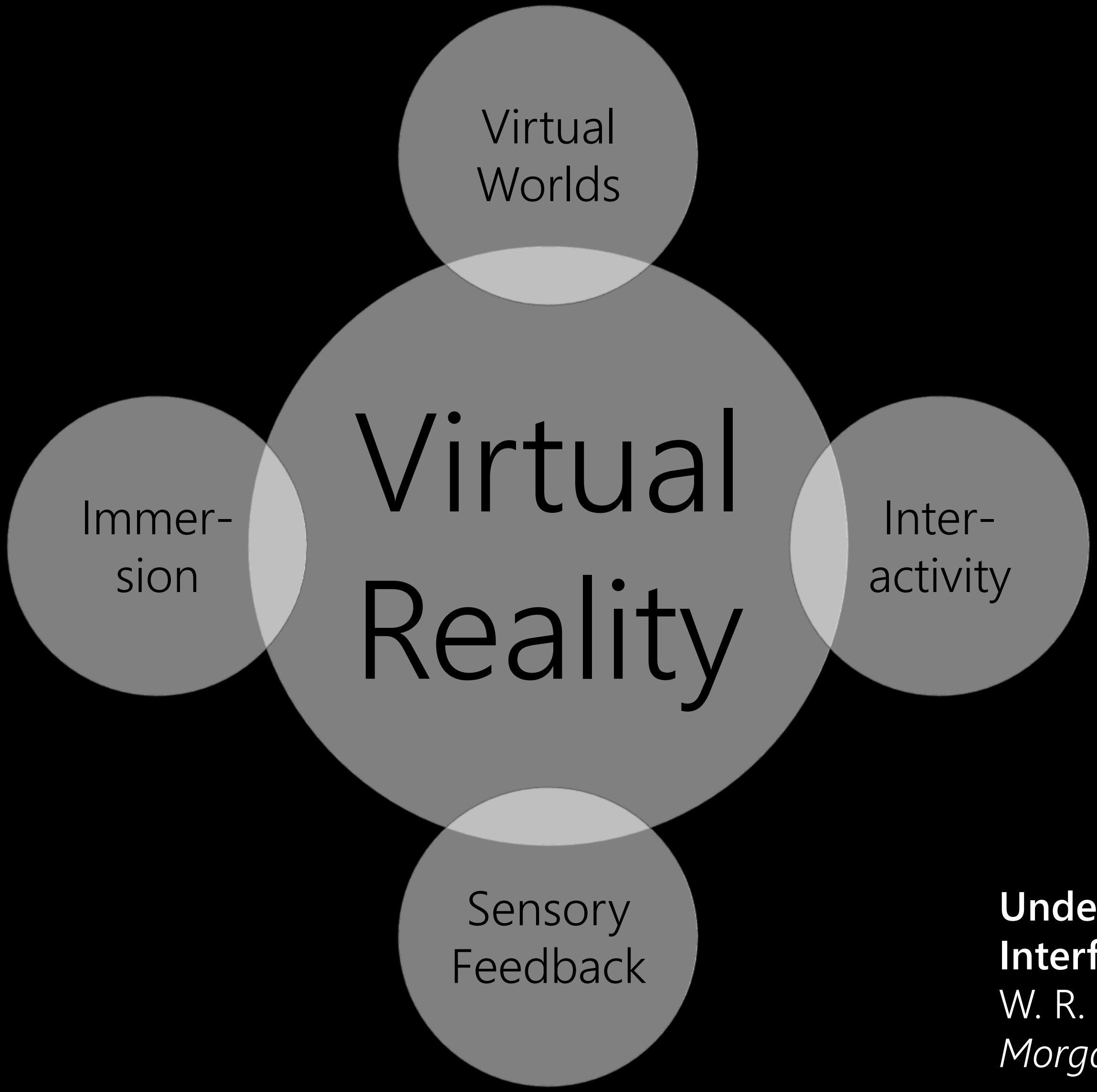
"The Sword of Damocles"
The world's first VR HMD, by Ivan Sutherland (1968)
Miniature CRTs, head tracking with mechanical sensors (in the video) or ultrasonic sensors

- Need to track motion to generate the right images:
  - head motion
  - hand motion
  - full-body motion
- Motion tracking enables:
  - **immersion** = the replacement of perception with virtual stimuli
  - **presence** = the sensation of "being there"

Motion-aware displays

1. Perception of immersion
2. Tracking in VR and AR
3. Hand input devices
4. Motion capture
5. Questions?

Virtual reality experiences



Understanding Virtual Reality: Interface, Application, and Design
W. R. Sherman & A. B. Craig
Morgan Kaufmann Publishers, 2003

Immersion vs Presence

- **Immersion** is an objective notion which can be defined as the sensory stimuli coming from a device, for example a data glove
- Measurable and comparable between devices

- **Presence** is a subjective phenomenon, personal experiences in an immersive environment
- Subjective feeling of being there

A note on presence terminology
M. Slater
Presence Connect, 2003, 3: 3

Slide adapted from Zerrin Yumak

Immersion

- sensation of being in another environment
- **Mental immersion:**
  - a movie, game or a novel might immerse you too
  - suspension of disbelief, state of being deeply engaged
- **Physical immersion:**
  - bodily entering into a medium
  - synthetic stimulus of the body's senses via the use of technology

Slide adapted from Zerrin Yumak

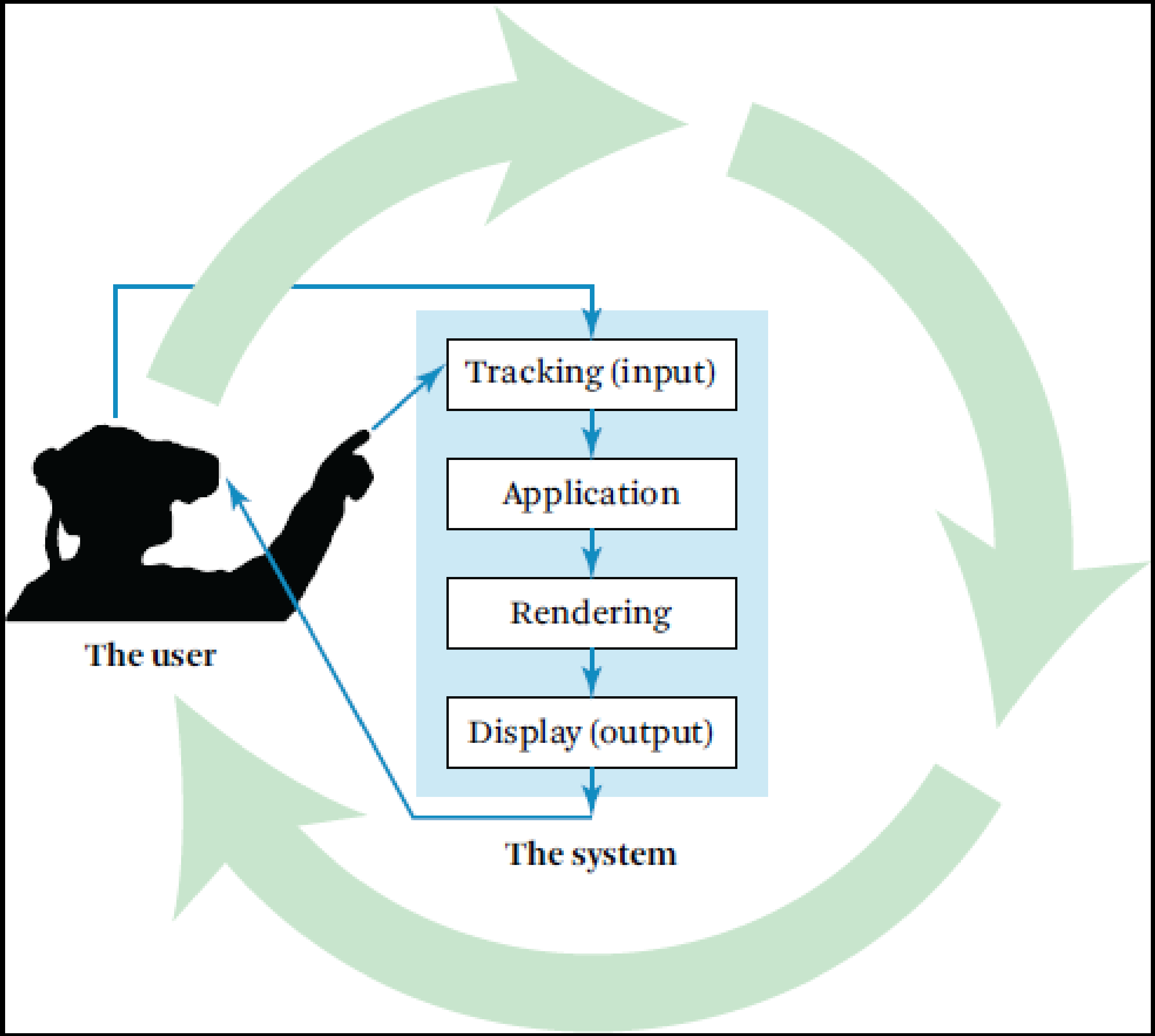
Self-embodiment

- Perception that the user has a body within the virtual world
- The presence of a virtual body can be quite compelling
  - even when that body does not look like one's own body
  - effective for teaching empathy by "walking in someone else's shoes" and can reduce racial bias
- Whereas body shape and colour are not so important, <u>motion is extremely important</u>
- Presence can be broken when visual body motion does not match physical motion

Putting Yourself in the Skin of a Black Avatar Reduces Implicit Racial Bias
T. C. Peck, S. Seinfeld, S. M. Aglioti & M. Slater
Consciousness and Cognition, 2013, 22(3), 779–787

Slide adapted from Zerrin Yumak

VR system input–output cycle

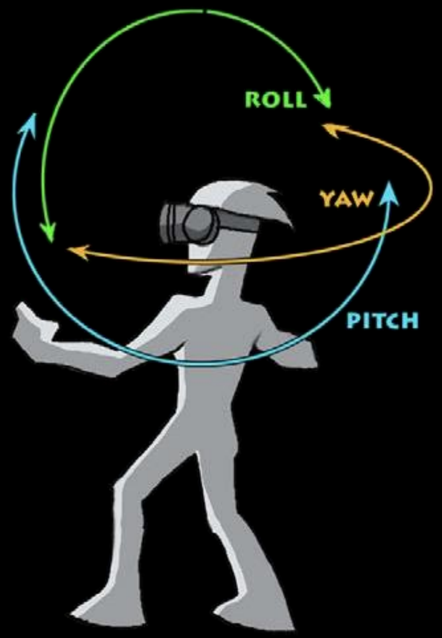


Scene-Motion- and Latency-Perception Thresholds for Head-Mounted Displays
J. J. Jerald
PhD Thesis, UNC Chapel Hill, 2009

Tracking degrees of freedom (DoF)

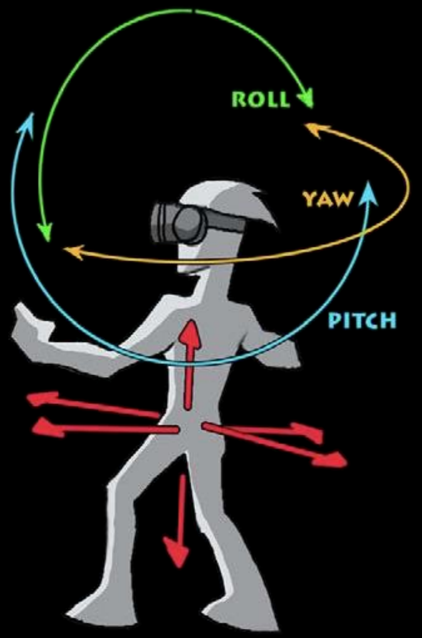
3 degrees of freedom (3-DoF)

- "In which direction am I looking"
- Detect rotational head movement
- Look around the virtual world from a fixed point



6 degrees of freedom (6-DoF)

- "Where am I and in which direction am I looking"
- Detect rotations and translational movement
- Move in the virtual world like in the real world



Slide adapted from Qualcomm Technologies, Inc.

Tracking technologies

- Mechanical:
  - e.g. physical linkage
- Electromagnetic:
  - e.g. magnetic sensing
- Inertial:
  - e.g. accelerometers, MEMs
- Acoustic:
  - e.g. ultrasonic
- Optical:
  - computer vision
- Hybrid:
  - combination of technologies

} contact-less tracking

Slide adapted from Bruce Thomas & Mark Billinghurst

Mechanical tracking

- Idea: mechanical arms with joint sensors
- Advantages:
  - high accuracy
  - low jitter
  - low latency
- Disadvantages:
  - cumbersome
  - limited range
  - fixed position

Ivan Sutherland (1968)

MicroScribe (2005)

Slide adapted from Bruce Thomas & Mark Billinghurst

Magnetic tracking

- Idea: measure difference in current between a magnetic transmitter and a receiver
- Advantages:
  - 6-DoF, robust & accurate
  - no line of sight needed
- Disadvantages:
  - limited range, noisy
  - sensible to metal
  - expensive

Razer Hydra (2011)
Magnetic source with two wired controllers
short range (<1 m), precision of 1 mm and 1°
62 Hz sampling rate, <50 ms latency

Slide adapted from Bruce Thomas & Mark Billinghurst

Inertial tracking

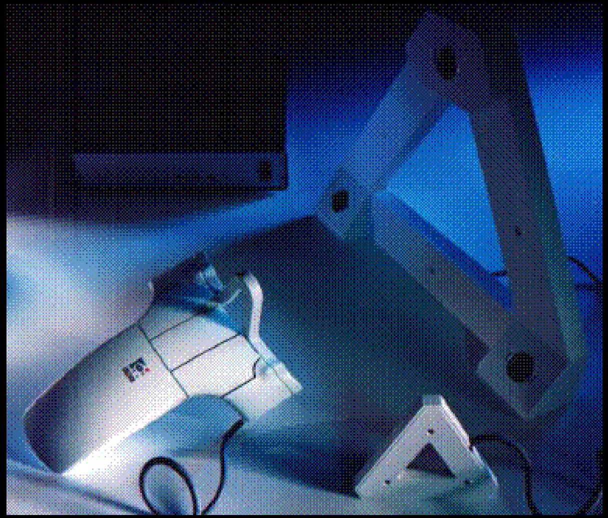
- Idea: Measuring linear and angular orientation rates (accelerometer/gyroscope)
- Advantages:
  - no transmitter, wireless
  - cheap + small
  - high sample rate
- Disadvantages:
  - drift + noise
  - only 3-DoF

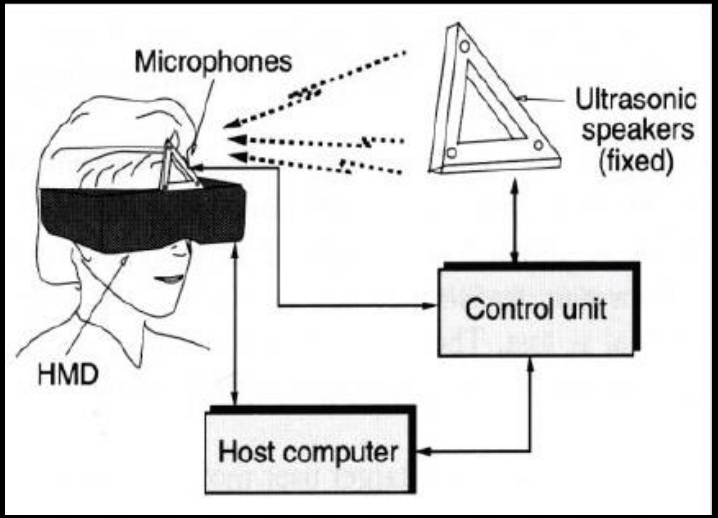
Google Daydream View (2017)
relies on the phone for processing and tracking
3-DoF rotational only tracking of phone + controller

Slide adapted from Bruce Thomas & Mark Billinghurst

Acoustic tracking

- Idea: time-of-flight or phase-coherent sound waves
- Advantages:
  - small + cheap
- Disadvantages:
  - only 3-DoF
  - low resolution
  - low sampling rate
  - requires line-of-sight
  - affected by environment (pressure, temperature)





Logitech 3D Head Tracker (1992)
Transmitter has 3 ultrasonic speakers, 30 cm apart; receiver has 3 mics
range: ~1.5 m, accuracy: 0.1° orientation, 2% distance
50 Hz update, 30 ms latency

Slide adapted from Bruce Thomas & Mark Billinghurst

Optical tracking

- Idea: image processing and computer vision to the rescue
- often using infrared light, retro-reflective markers, multiple views
- Advantages:
  - long range, cheap
  - immune to metal
  - usually very accurate
- Disadvantages:
  - requires markers, line of sight
  - can have low sampling rate

Microsoft Kinect (2010)
IR laser speckle projector, RGB + IR cameras
range: 1–6 m, accuracy: <5 mm
30 Hz update rate, 100 ms latency

Slide adapted from Bruce Thomas & Mark Billinghurst

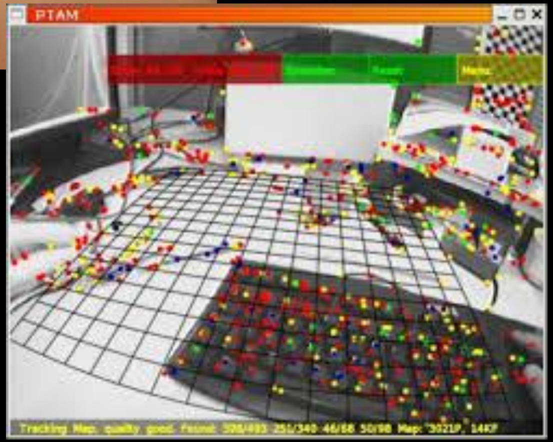
AR optical tracking

- Marker tracking:
  - tracking known artificial images
    - e.g. ARToolKit square markers
- Markerless tracking:
  - tracking from known features in real world
    - e.g. Vuforia image tracking
- Unprepared tracking:
  - in unknown environments
    - e.g. SLAM (simultaneous localisation and mapping)

devfun-lab.com

mobilegeeks.de

PTAM



Slide adapted from Bruce Thomas & Mark Billinghurst

Hybrid tracking

- Idea: multiple technologies overcome limitations of each one
- A system that utilizes two or more position/orientation measurement technologies (e.g. inertial + visual)
- Advantages:
  - robust
  - reduce latency
  - increase accuracy
- Disadvantages:
  - more complex + expensive

digitaltrends.com

Apple ARKit (2017), Google ARCore (2018)
visual-inertial odometry – combine inertial motion sensing with feature point tracking

Slide adapted from Bruce Thomas & Mark Billinghurst

Example: Vive Lighthouse tracking

- Outside-in hybrid tracking:
  - 2 base stations: each with 2 laser scanners, LED array
- Headworn/handheld sensors:
  - 37 photo sensors in HMD, 17 in hand
  - additional IMU sensors (500 Hz)
- Performance:
  - tracking fuses sensor samples at 250 Hz
  - 2 mm RMS accuracy
  - large area: 5×5 m² range
- See: https://youtu.be/xrsUMEbLtOs

gizmodo.com

slashgear.com

Slide adapted from Bruce Thomas & Mark Billinghurst

Hand input devices

- Devices that integrate hand input into VR:
  - world-grounded input devices
  - non-tracked handheld controllers
  - tracked handheld controllers
  - hand-worn devices
  - hand tracking

digitaltrends.com

Slide adapted from Bruce Thomas & Mark Billinghurst

World-grounded hand input devices

- Devices constrained or fixed in the real world
  - e.g. joysticks, steering wheels
- Not ideal for VR
  - constrains user motion
- Good for VR vehicle metaphor, location-based entertainment
  - e.g. driving simulators, Disney's "Aladdin's Magic Carpet Ride"

aliexpress.com

realityprime.com

Slide adapted from Bruce Thomas & Mark Billinghurst

Non-tracked handheld controllers

- Devices held in hand
  - buttons
  - joysticks
  - game controllers
- Traditional video game controllers
  - e.g. Xbox controller

Bottomless Joystick
katsumotoy.com/bj/

techadvisor.co.uk

Slide adapted from Bruce Thomas & Mark Billinghurst

Tracked handheld controllers

- Handheld controller with 6-DoF tracking
  - combines button/joystick/ trackpad input plus tracking
- One of the best options for VR applications
  - physical prop enhancing VR presence
  - providing proprioceptive, passive haptic touch cues
  - direct mapping to real hand motion

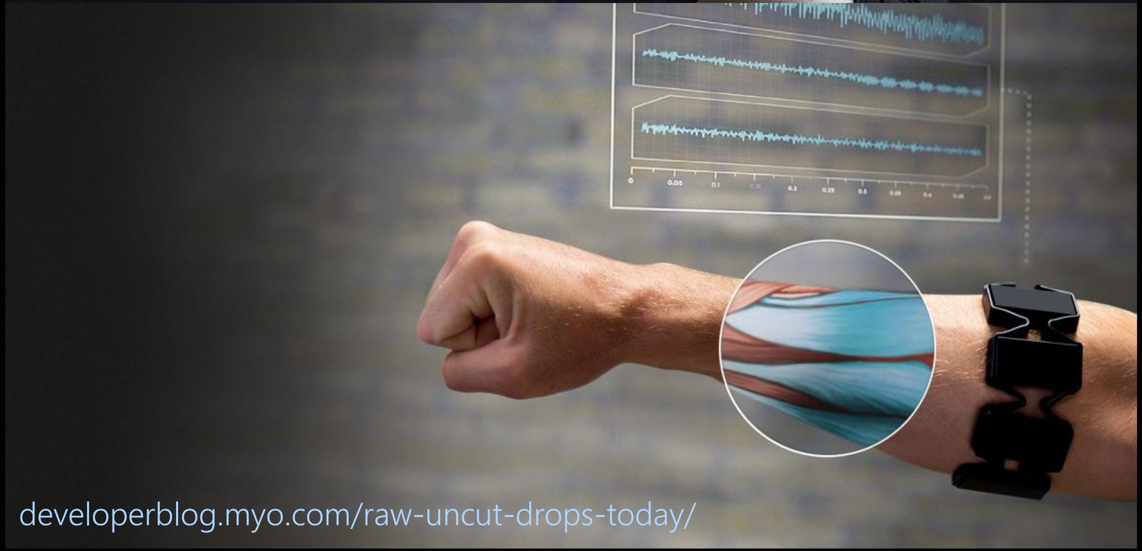
HTC Vive controller

Oculus Touch

Slide adapted from Bruce Thomas & Mark Billinghurst

Hand-worn devices

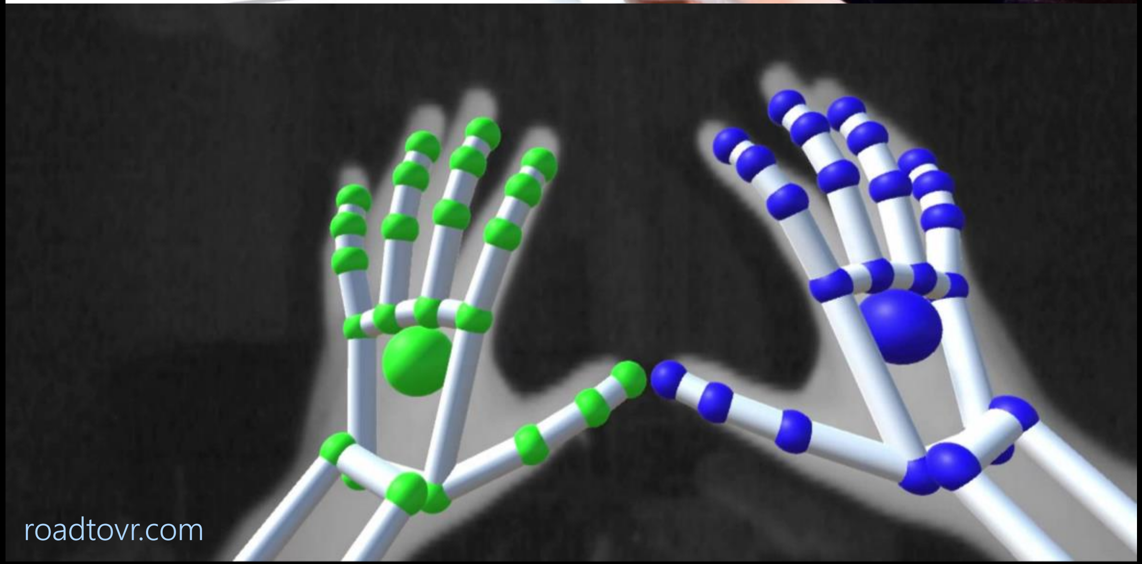
- Devices worn on hands/arms
  - e.g. glove, EMG sensors, rings
- Advantages:
  - natural input with potentially rich gesture interaction
  - hands can be held in comfortable positions
    - no line-of-sight issues
  - hands and fingers can fully interact with real objects



Slide adapted from Bruce Thomas & Mark Billinghurst

Hand tracking

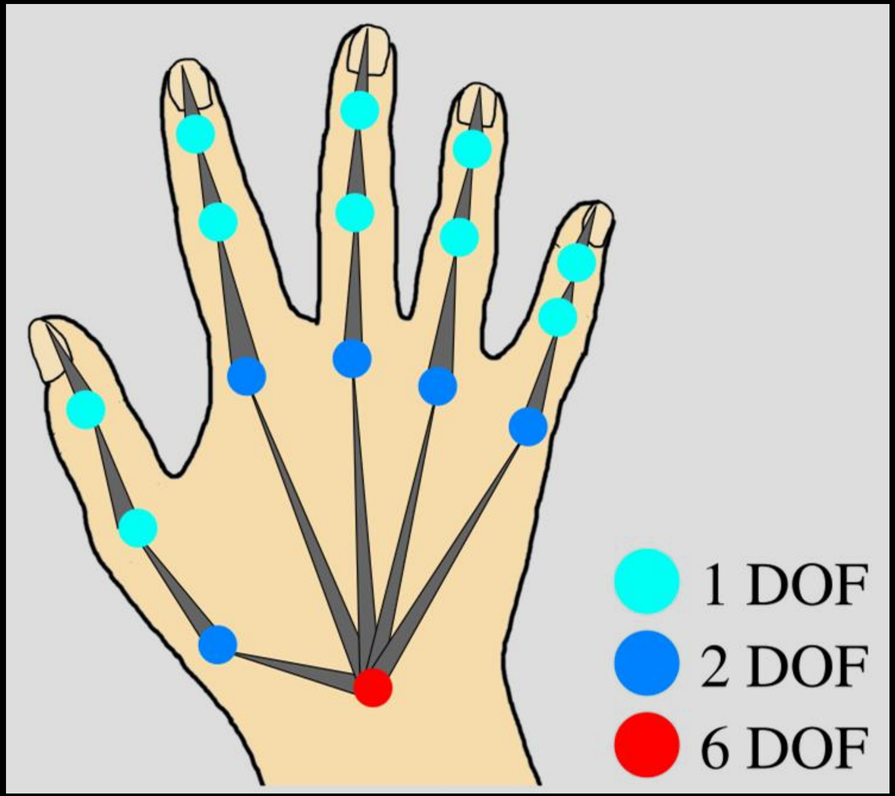
- Using computer vision to track bare hand input
- Creates compelling sense of presence, natural interaction
- Advantages:
  - least intrusive, purely passive
  - hands-free tracking, so can interact freely with real objects
  - low power requirements, cheap
  - more ubiquitous, works outdoors



Slide adapted from Bruce Thomas & Mark Billinghurst, Fransizka Mueller

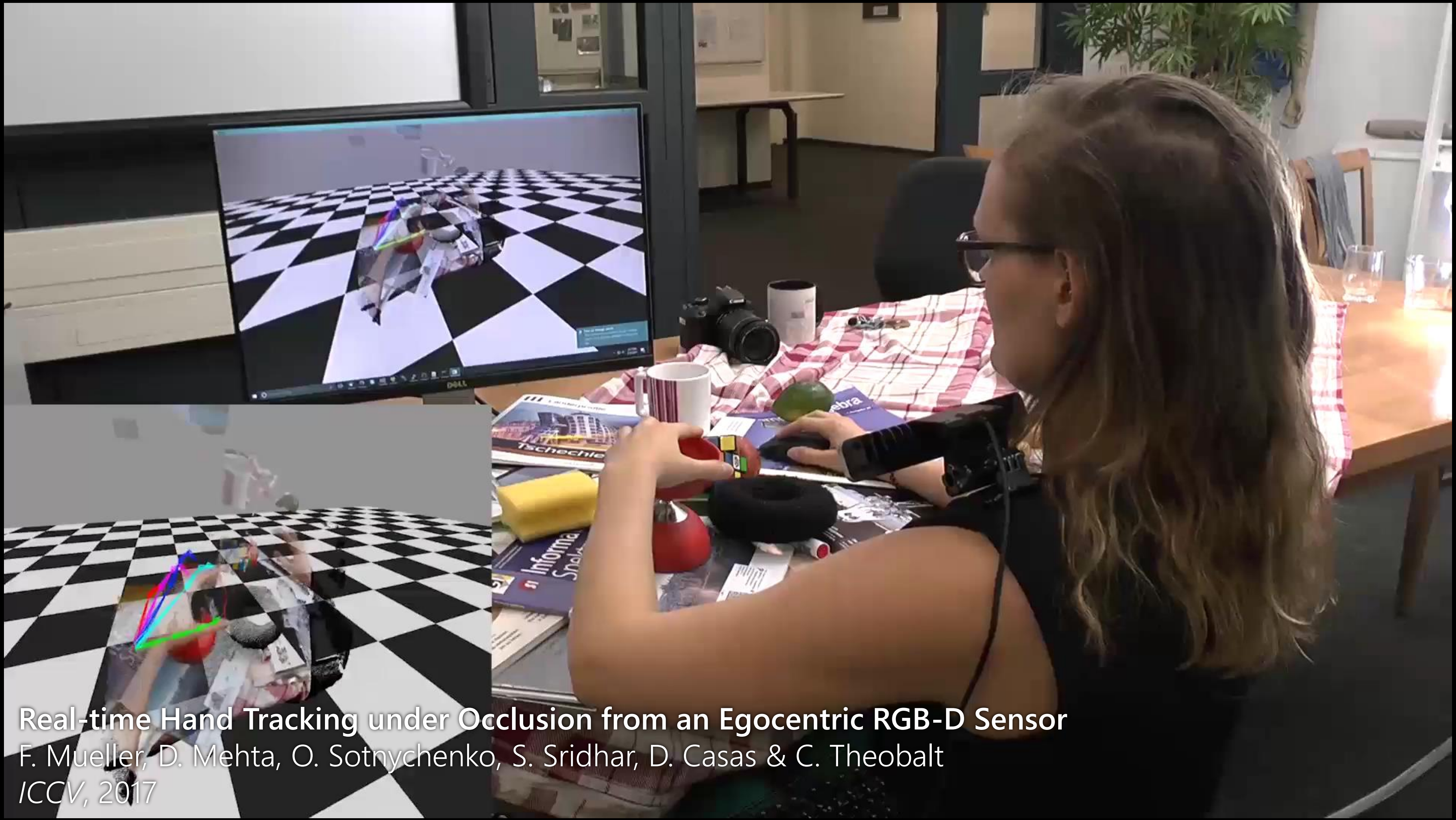
Case study: Egocentric hand tracking

- **Goal:** reconstruct full hand pose (global transform + joint angles) using a single body-mounted camera
- Robust to:
  - fast and complex motions
  - background clutter
  - occlusions by arbitrary objects as well as the hand itself
  - self-similarities of hands
  - fairly uniform colour
- In real time (>30 Hz)



Slide adapted from Fransizka Mueller

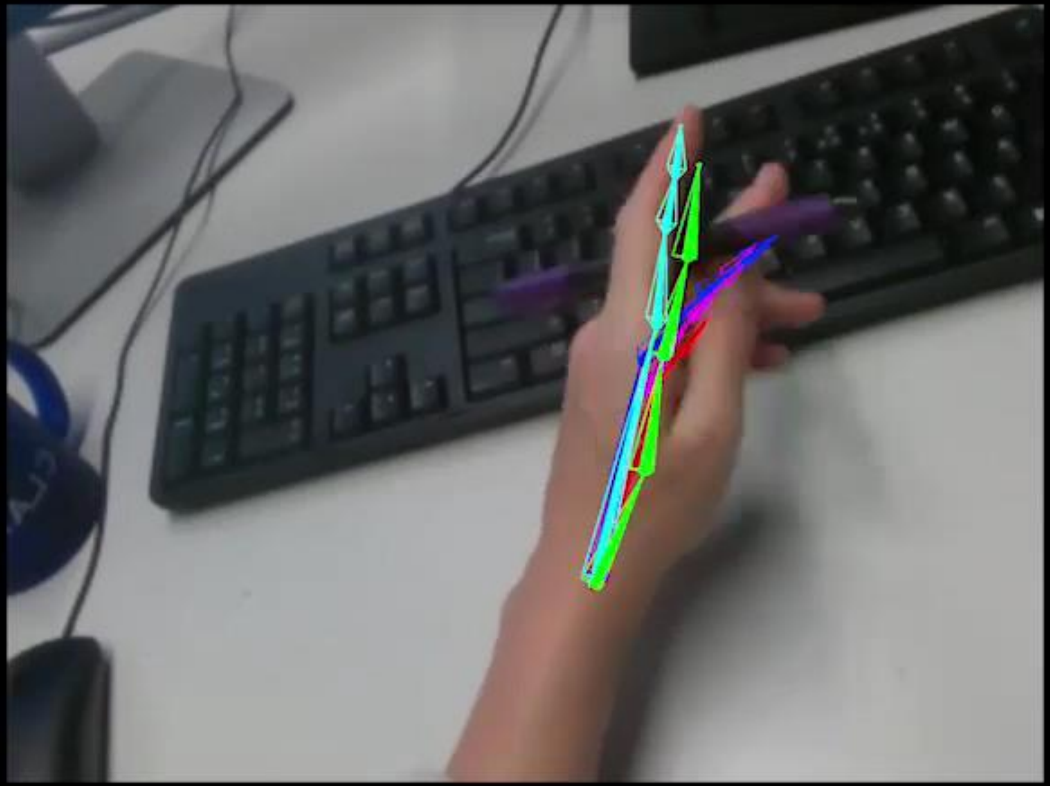
Egocentric hand tracking from RGB-D

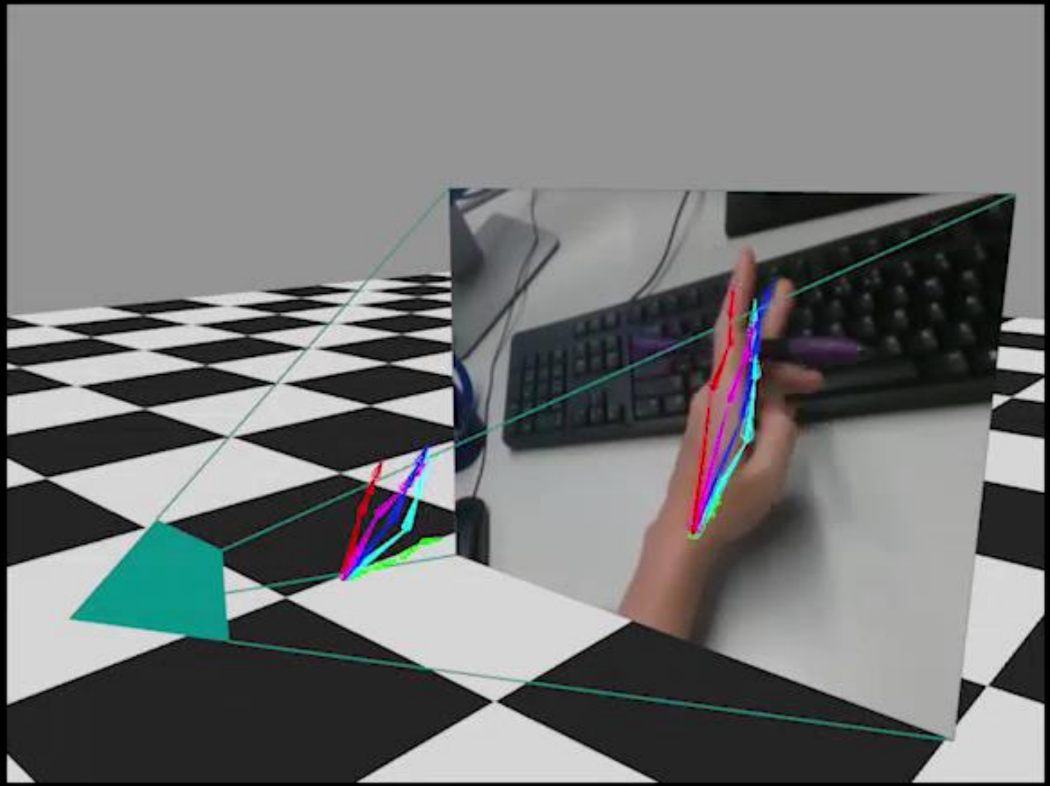


Real-time Hand Tracking under Occlusion from an Egocentric RGB-D Sensor
F. Mueller, D. Mehta, O. Sotnychenko, S. Sridhar, D. Casas & C. Theobalt
ICCV, 2017

https://youtu.be/Aay3MgHqQ0k?t=296

Egocentric hand tracking



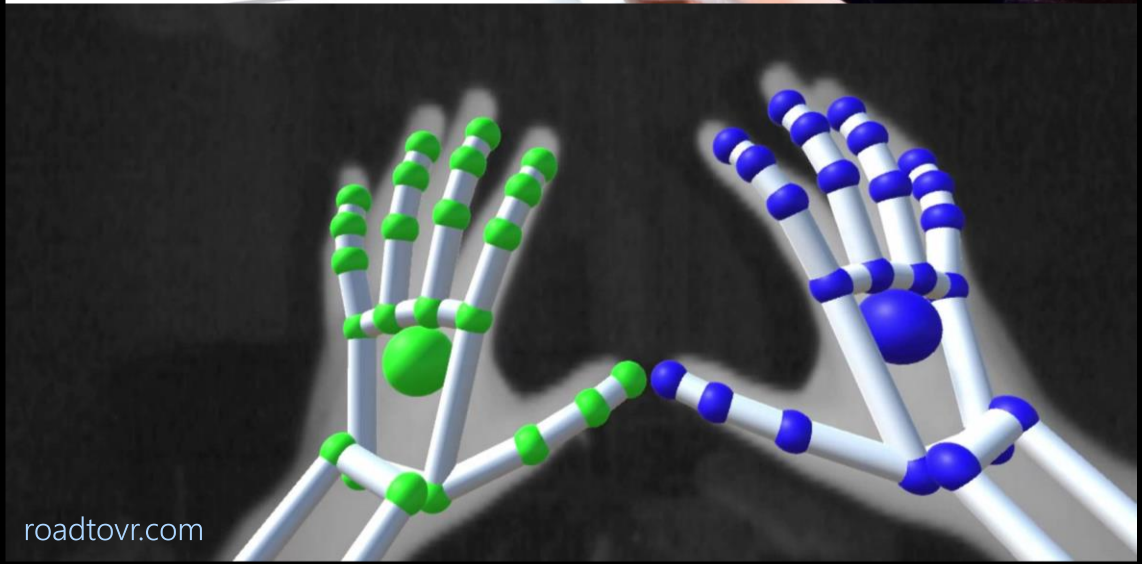


https://youtu.be/0wH0b9MdjPI?t=4

GANerated Hands for Real-time 3D Hand Tracking from Monocular RGB
F. Mueller, F. Bernard, O. Sotnychenko, D. Mehta, S. Sridhar, D. Casas & C. Theobalt
CVPR, 2018

Remaining challenges of hand tracking

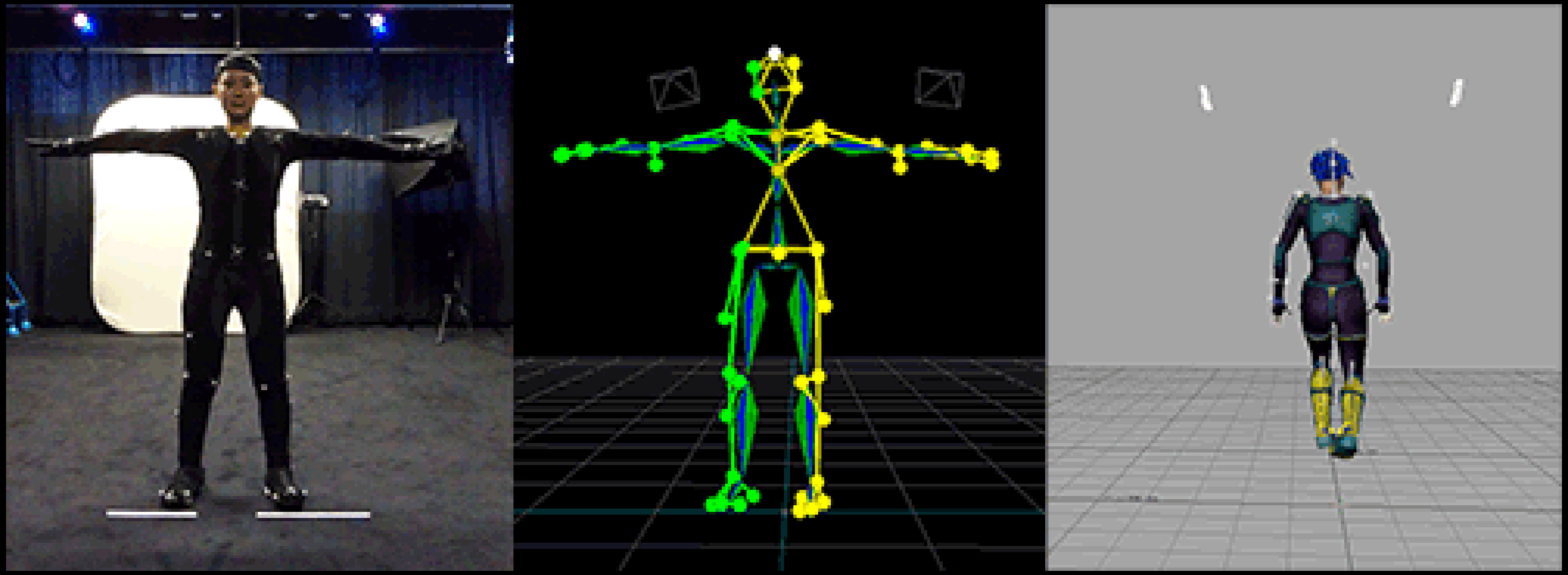
- Robust results out of the box:
  - interacting with unknown objects
  - two hands simultaneously
  - no explicit model fitting
- Usability challenges:
  - not having sense of touch
  - line of sight required to sensor
  - fatigue from holding hands in front of sensor



Slide adapted from Bruce Thomas & Mark Billinghurst

Full-body tracking

- Adding full-body input into VR:
  - creates illusion of self-embodiment
  - significantly enhances sense of presence



Slide adapted from Bruce Thomas & Mark Billinghurst

Camera-based motion capture

- Use multiple cameras (8+) with infrared (IR) LEDs
- Retro-reflective markers on body clearly reflect IR light
- For example Vicon, OptiTrack:
  - very accurate: <1 mm error
  - very fast:
    - 100–360 Hz sampling rate
    - <10 ms latency
  - each marker needs to be seen by at least two cameras

Vicon

OptiTrack

digitalcinema.com.ua

Slide adapted from Bruce Thomas & Mark Billinghurst

SIGGRAPH ASIA 2016 MACAO

EgoCap: Egocentric Marker-less Motion Capture with Two Fisheye Cameras

Helge Rhodin[1] Christian Richardt[123] Dan Casas[1],
Eldar Insafutdinov[1] Mohammad Shafiei[1]
Hans-Peter Seidel[1] Bernt Schiele[1] Christian Theobalt[1]

Graphics Vision & Video [1] mpii max planck institut informatik [2] intel Visual Computing Institute [3] University of Bath

Today's motion-capture challenges

- General environments
- Large scale motions
- Constrained rooms
- Easy to use, non-intrusive
- Low delay

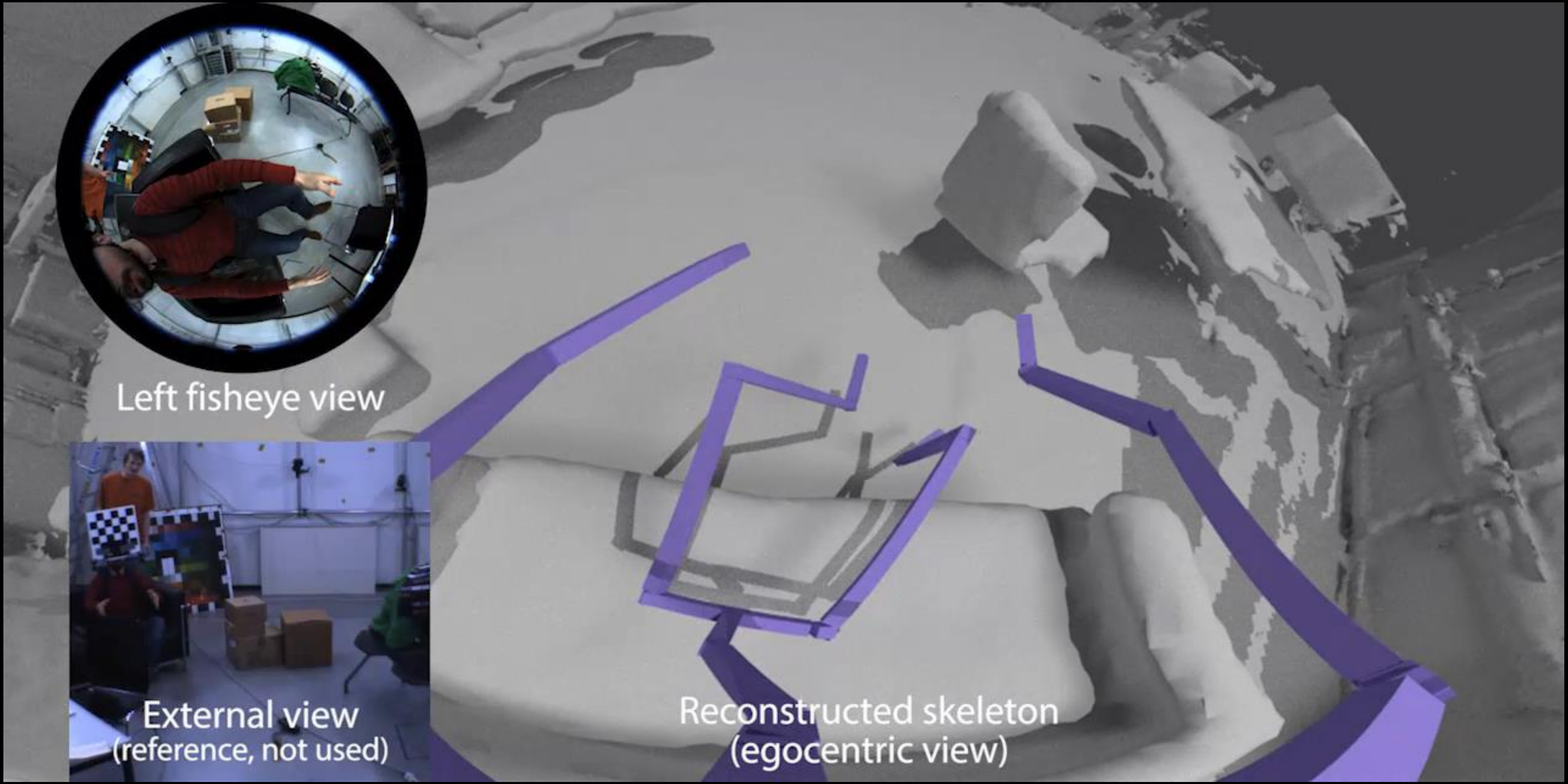
Computer animation

Sports and medicine

Autonomous driving

Virtual and augmented reality

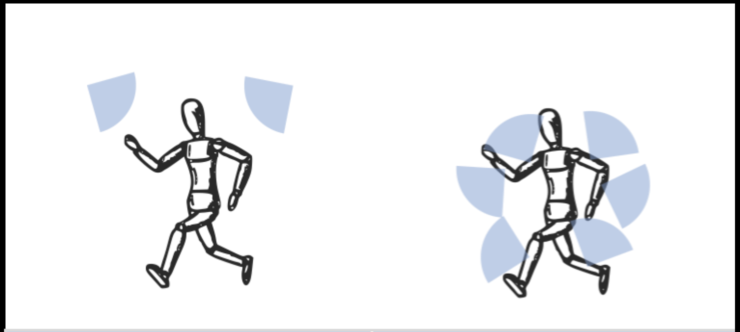
Embodied virtual reality



Marker-less motion capture

Outside-in
Non-intrusive
Limited capture volume
Full-body

Marker-less motion capture



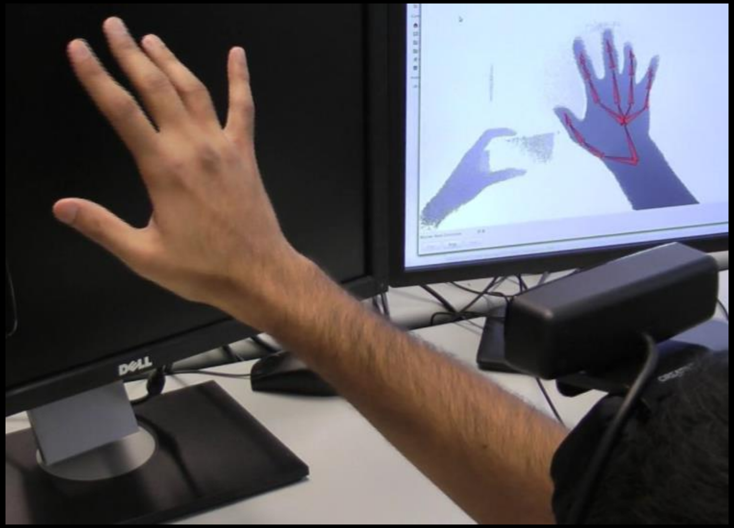
Outside-in	Inside-out
Non-intrusive	Intrusive
Limited capture volume	Infinite capture volume
Full-body	Full-body

[Shiratori 2011]

Marker-less motion capture

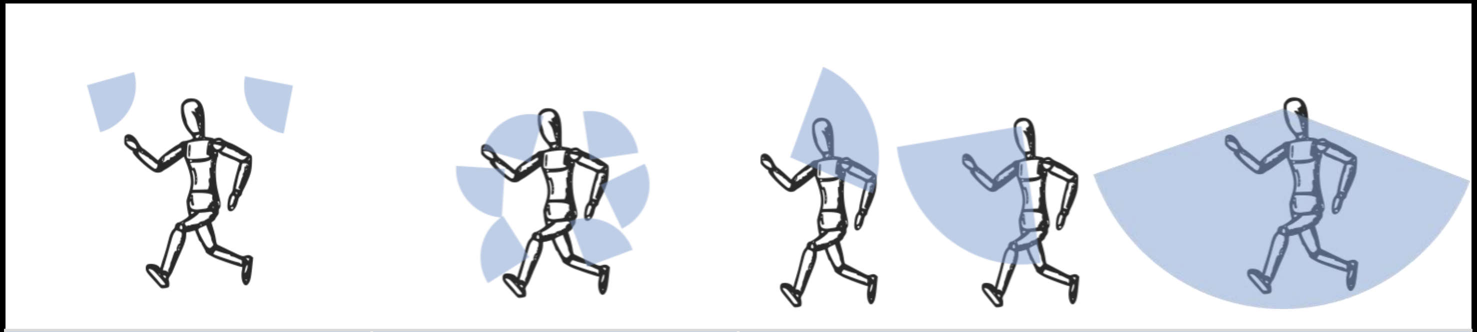
Outside-in	Inside-out	Inside-in
Non-intrusive	Intrusive	Low intrusion
Limited capture volume	Infinite capture volume	Infinite capture volume
Full-body	Full-body	Partial-body

[Jones 2011, Wang 2016]



[Sridhar 2015, ...]

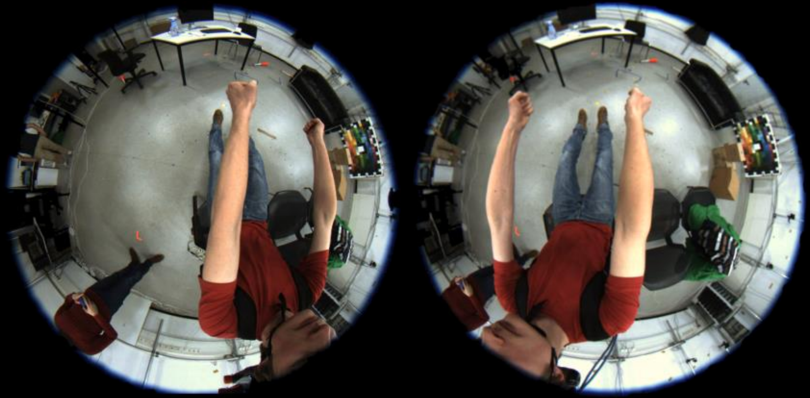
Marker-less motion capture



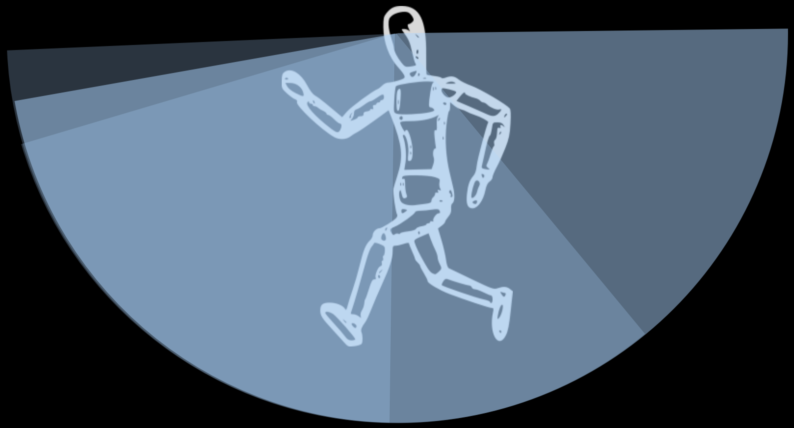
Outside-in	Inside-out	Inside-in	EgoCap
Non-intrusive	Intrusive	Low intrusion	Low intrusion
Limited capture volume	Infinite capture volume	Infinite capture volume	Infinite capture volume
Full-body	Full-body	Partial-body	Full-body

Camera gear

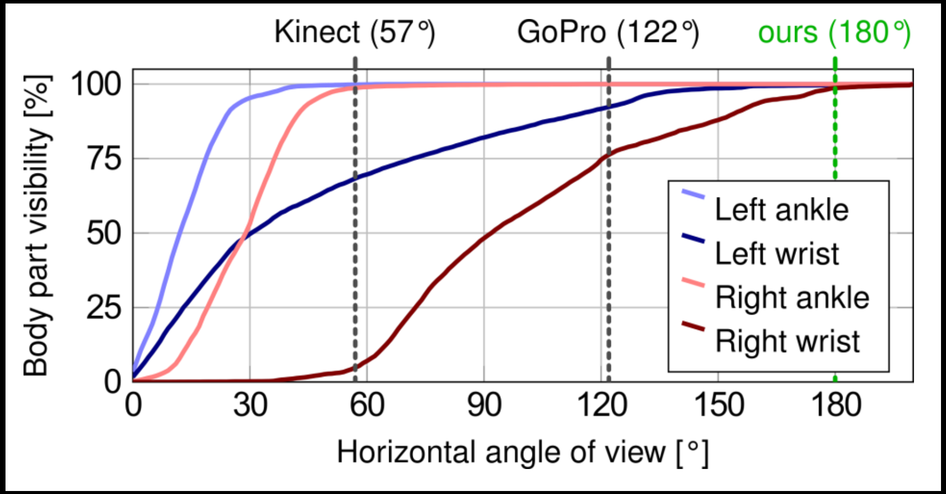
Camera extensions



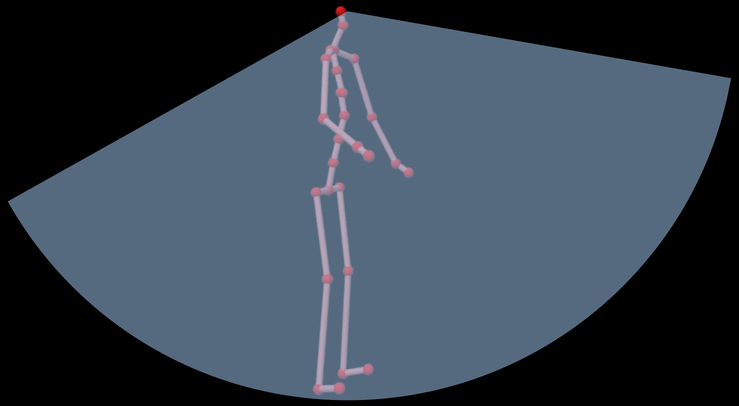
Egocentric view examples



Field of view



Egocentric capture challenges



Camera is attached

Subject is always in view

Human pose is independent of global motion

Estimation of global motion

Moving background

Top-down view

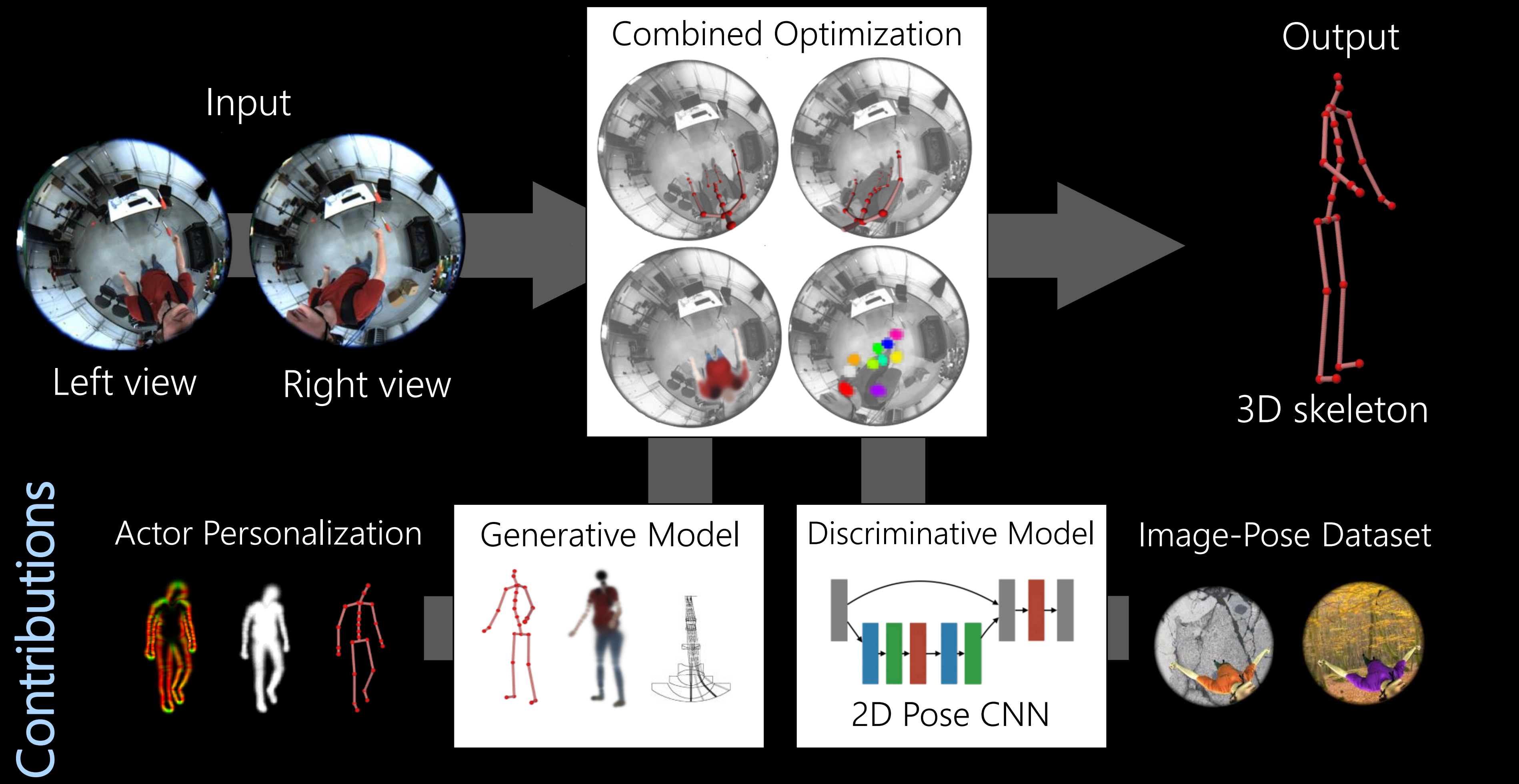
Self-occlusions

The lower body appears tiny

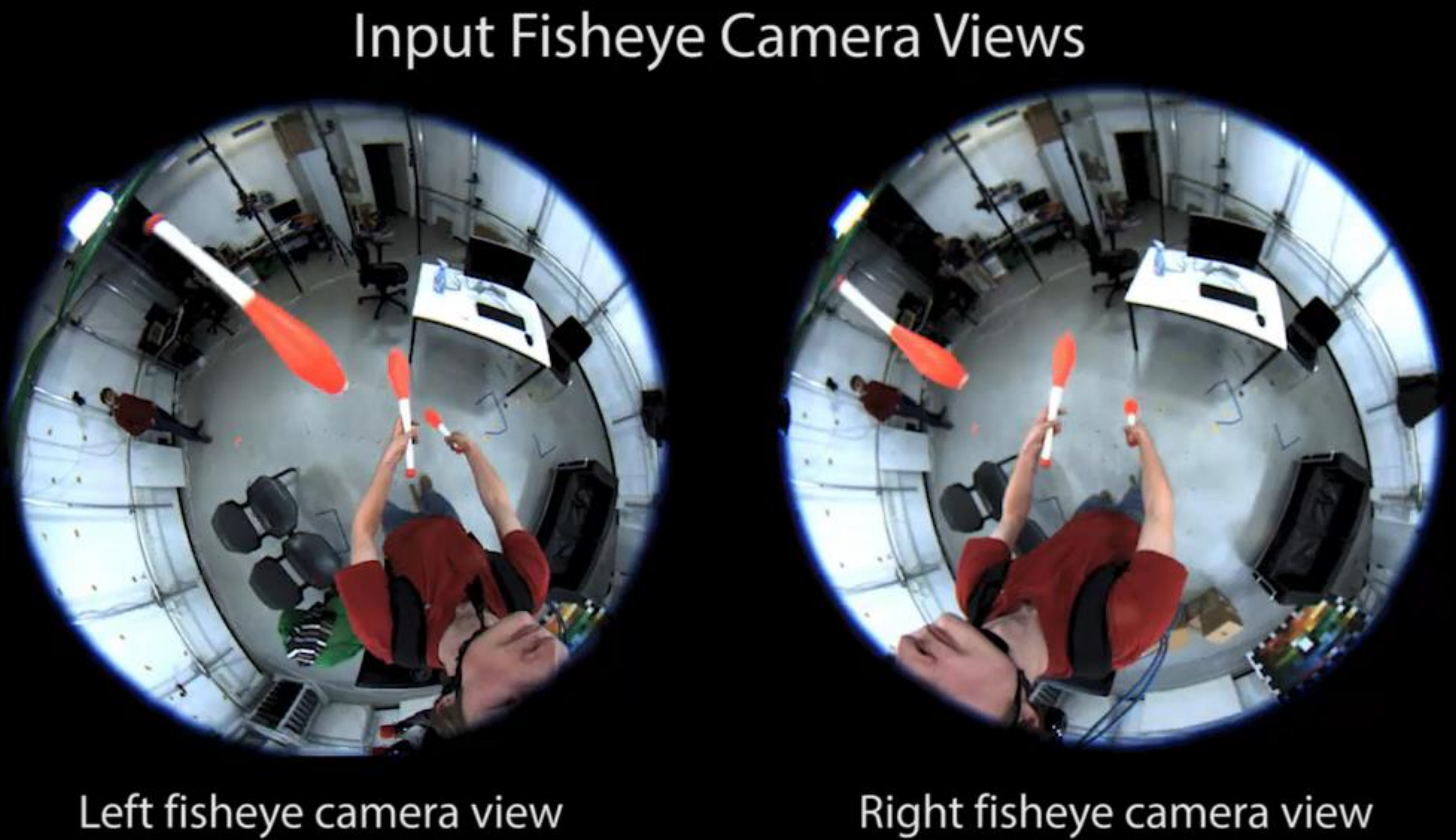
RGB only

Depth ambiguities

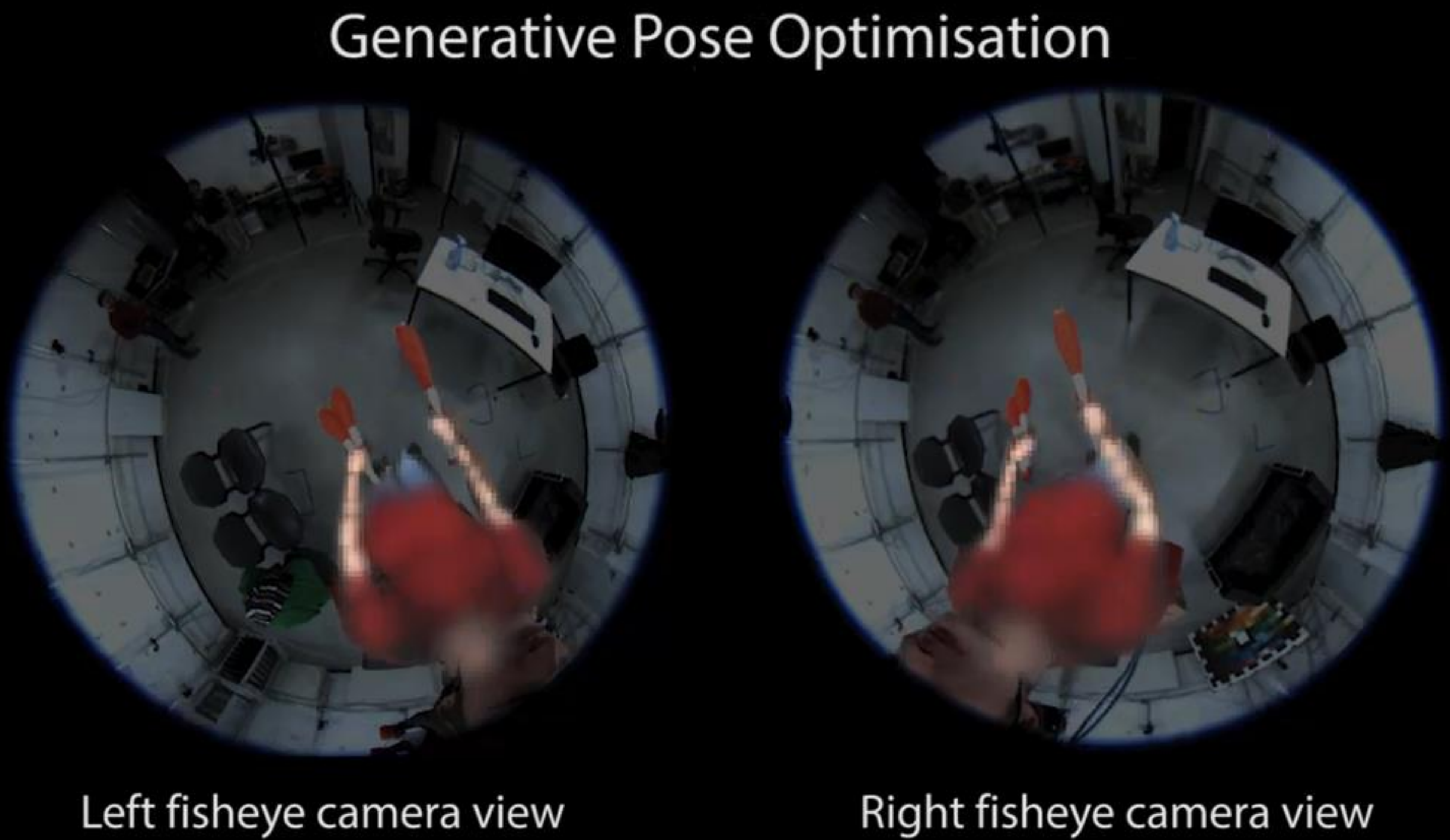
Model overview



Method walkthrough



Method walkthrough

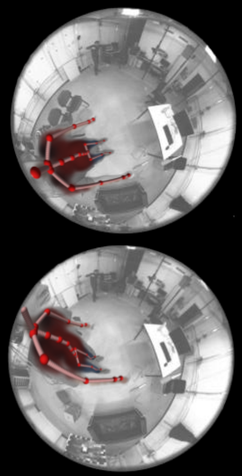


Combined optimization

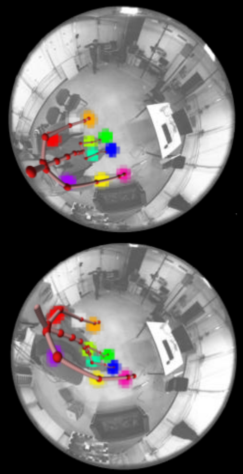
- Energy minimization:
  - gradient descent on pose $\mathbf{p}^t$ at time t

$$E(\mathbf{p}^t) = E_{\text{color}}(\mathbf{p}^t) + E_{\text{detection}}(\mathbf{p}^t) + E_{\text{pose}}(\mathbf{p}^t) + E_{\text{smooth}}(\mathbf{p}^t)$$

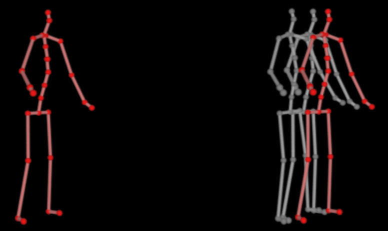
Input



Generative



Discriminative



Prior terms

Importance of energy terms

Without body-part detection term (Section 4.3.3)

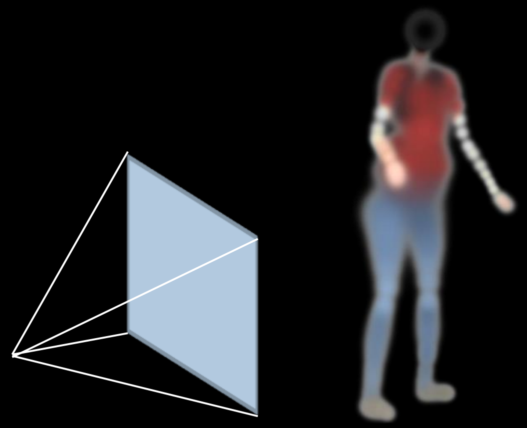
Importance of energy terms

Complete energy

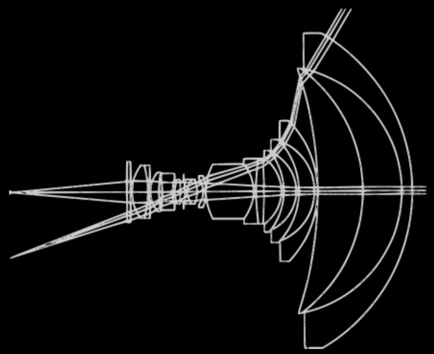
Complete energy (with smoothing)

Generative model

- Volumetric body model
  - raytracing-based
  - fisheye camera
  - parallel GPU implementation



[Rhodin ICCV 2015, ECCV 2016]



[Scaramuzza 2006]

Our model

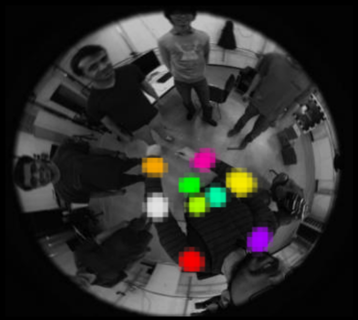
Discriminative component

- Deep 2D pose estimation
  - High accuracy with sufficient training data
  - Standard CNN architecture (Residual network [He 2016])

[Insafutdinov 2016, …]

- Egocentric training data?

Example image

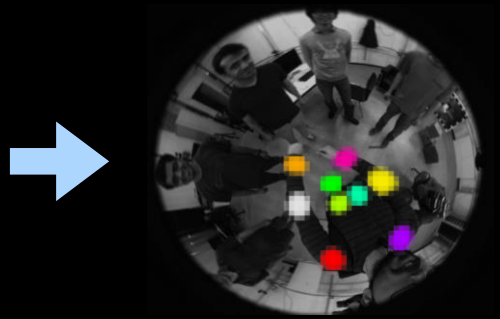


Annotation

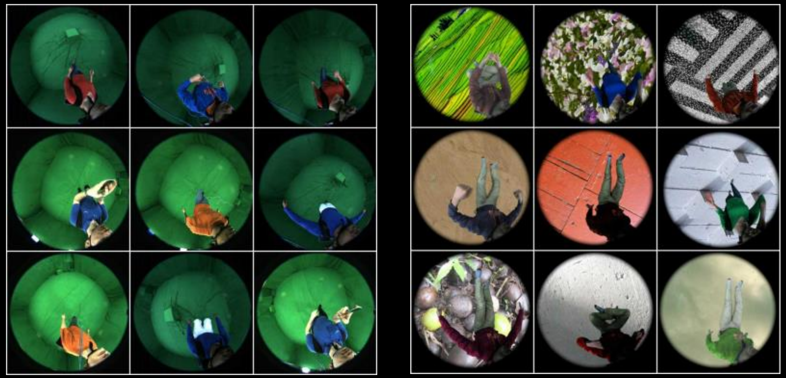
Training dataset

- Egocentric image-pose database
  - 80,000 images
  - appearance variation
  - background variation
  - actor variation

Example image



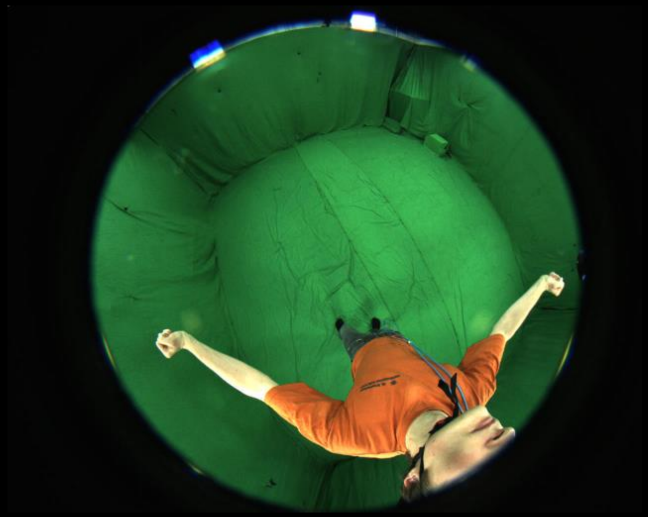
Annotation



Data augmentation

Ground-truth annotation

Diversity by augmentation: background



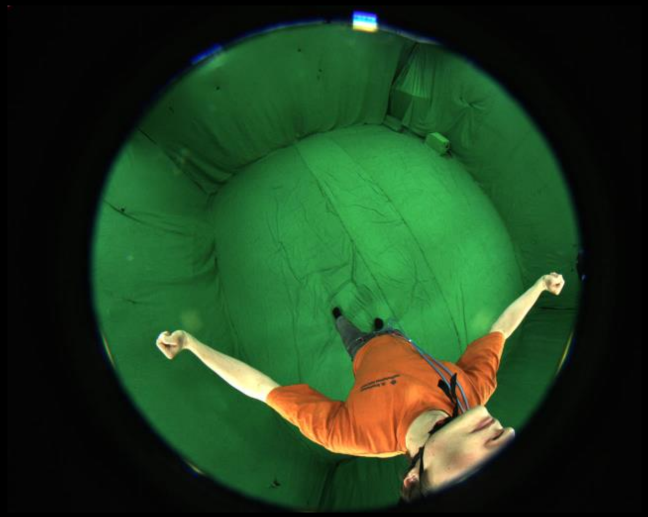
Original

Augmentation

Replaced background

- Green-screen keying to replace backgrounds
  - using random images from Flickr

Diversity by augmentation: foreground

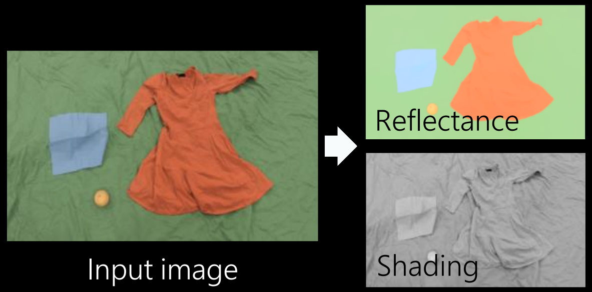


Original

Augmentation

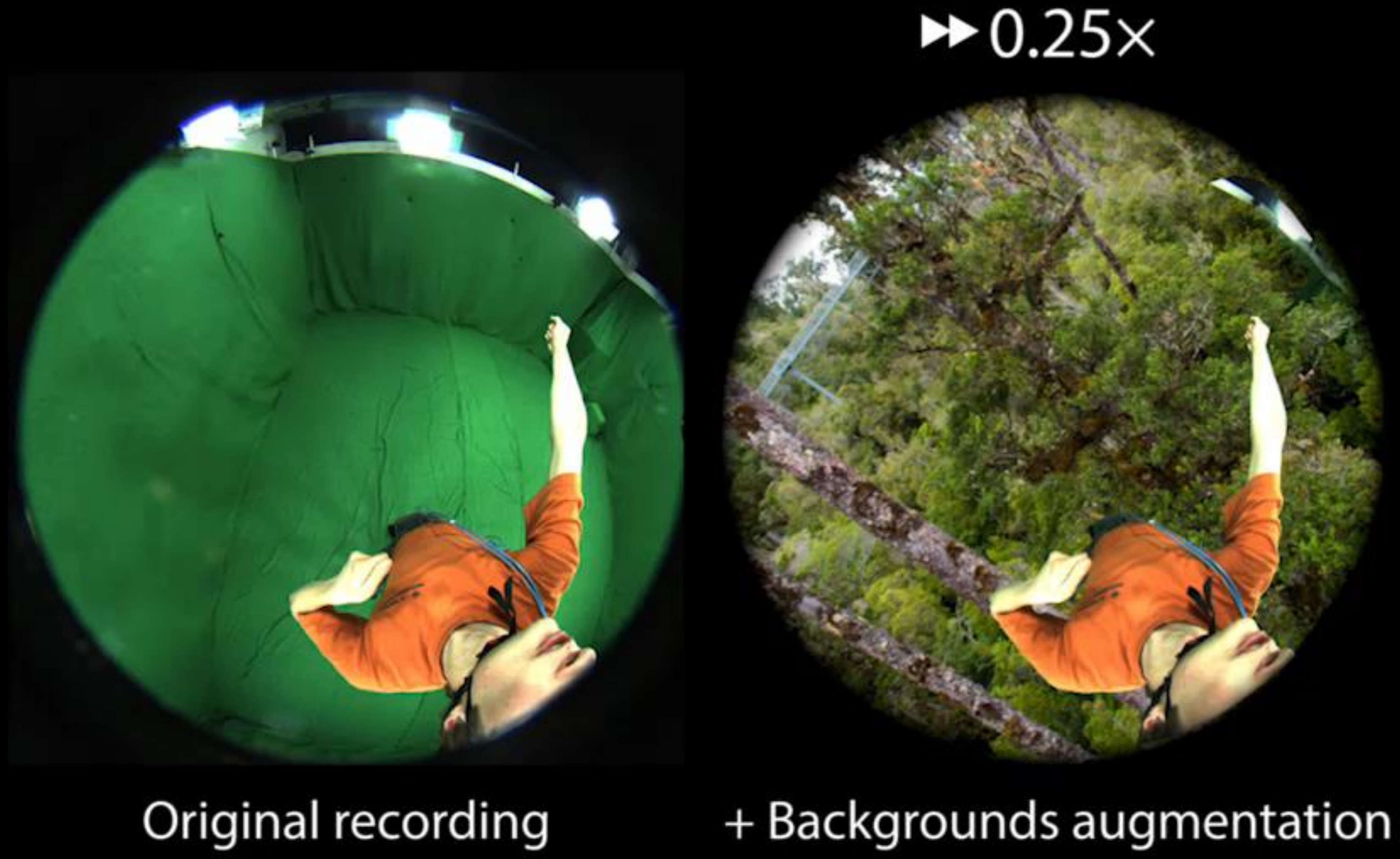
Replaced albedo

- Intrinsic image decomposition [Meka 2016, …]



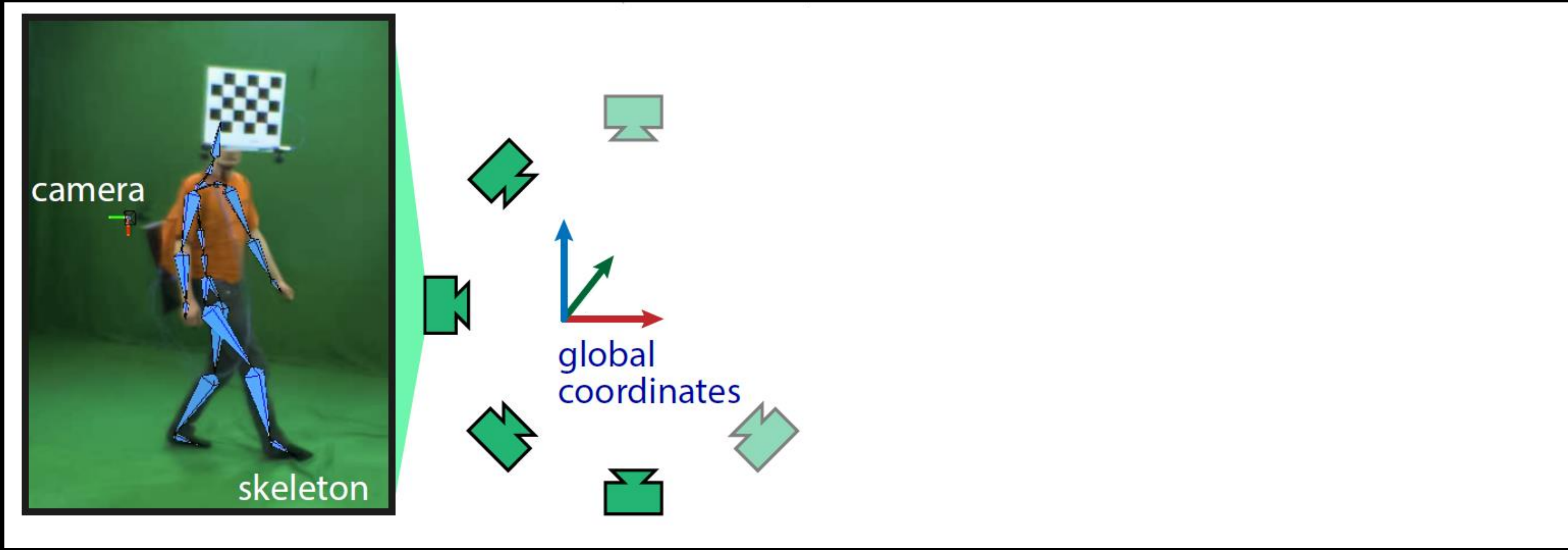
Input image

Training dataset augmentation



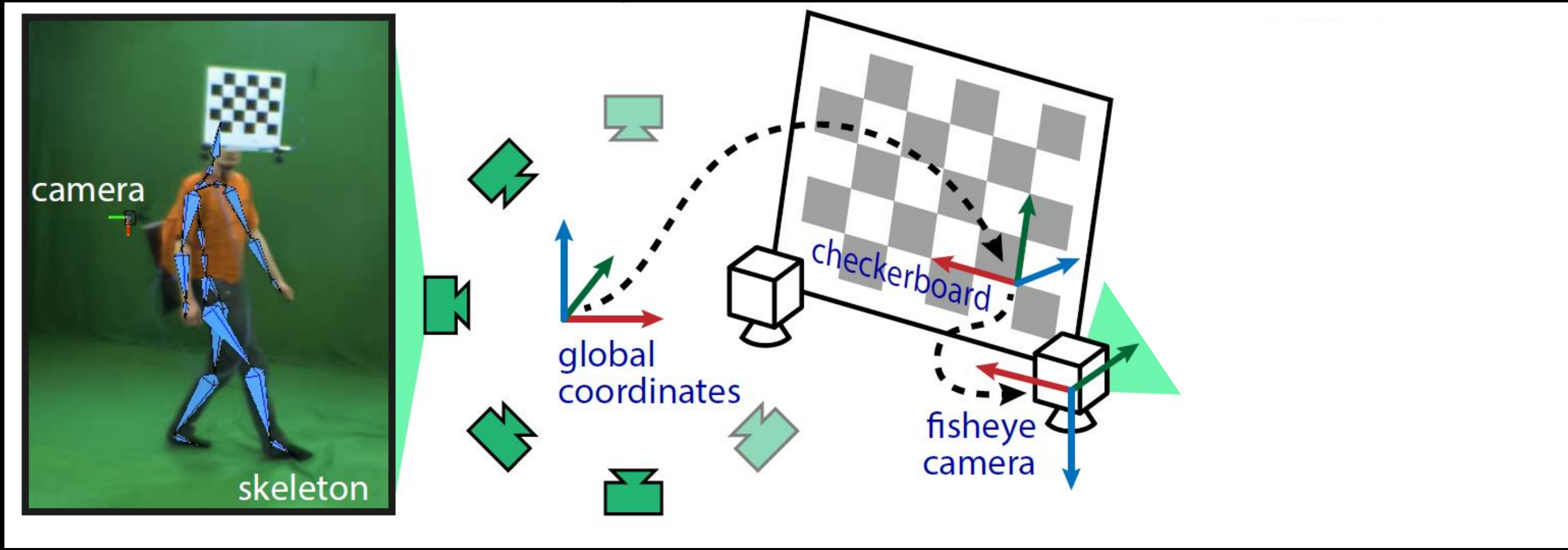
Automatic ground-truth annotation

Outside-in markerless motion capture



Automatic ground-truth annotation

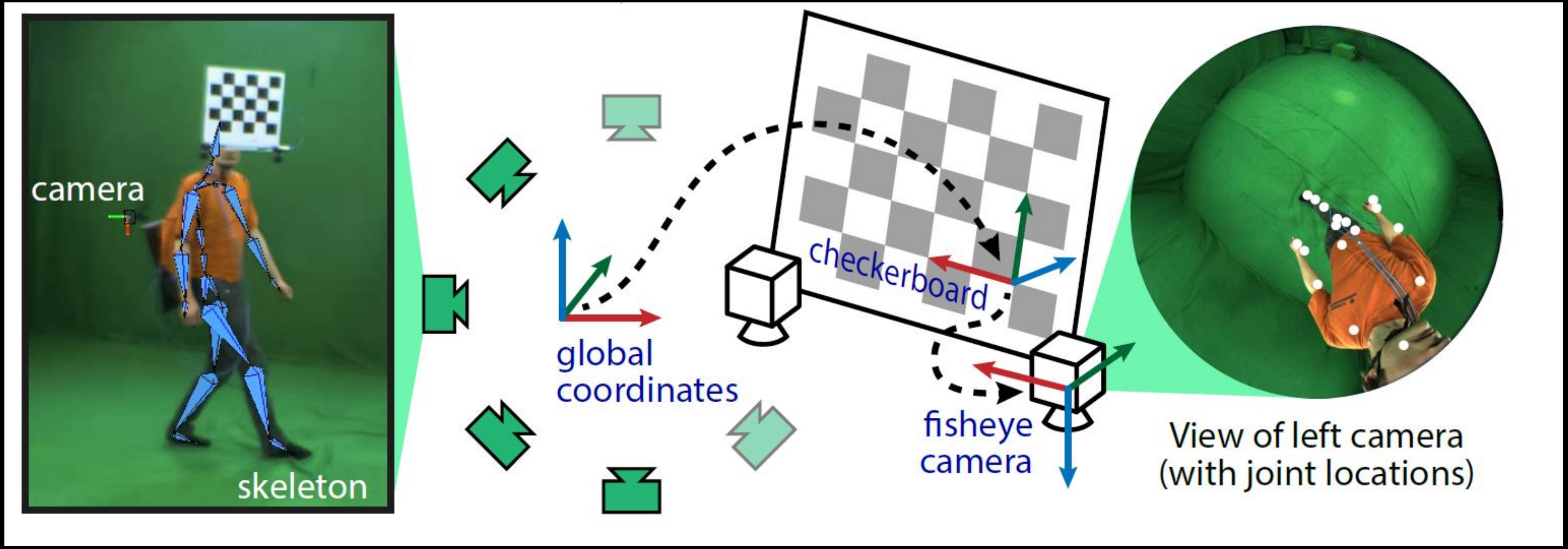
Outside-in markerless motion capture



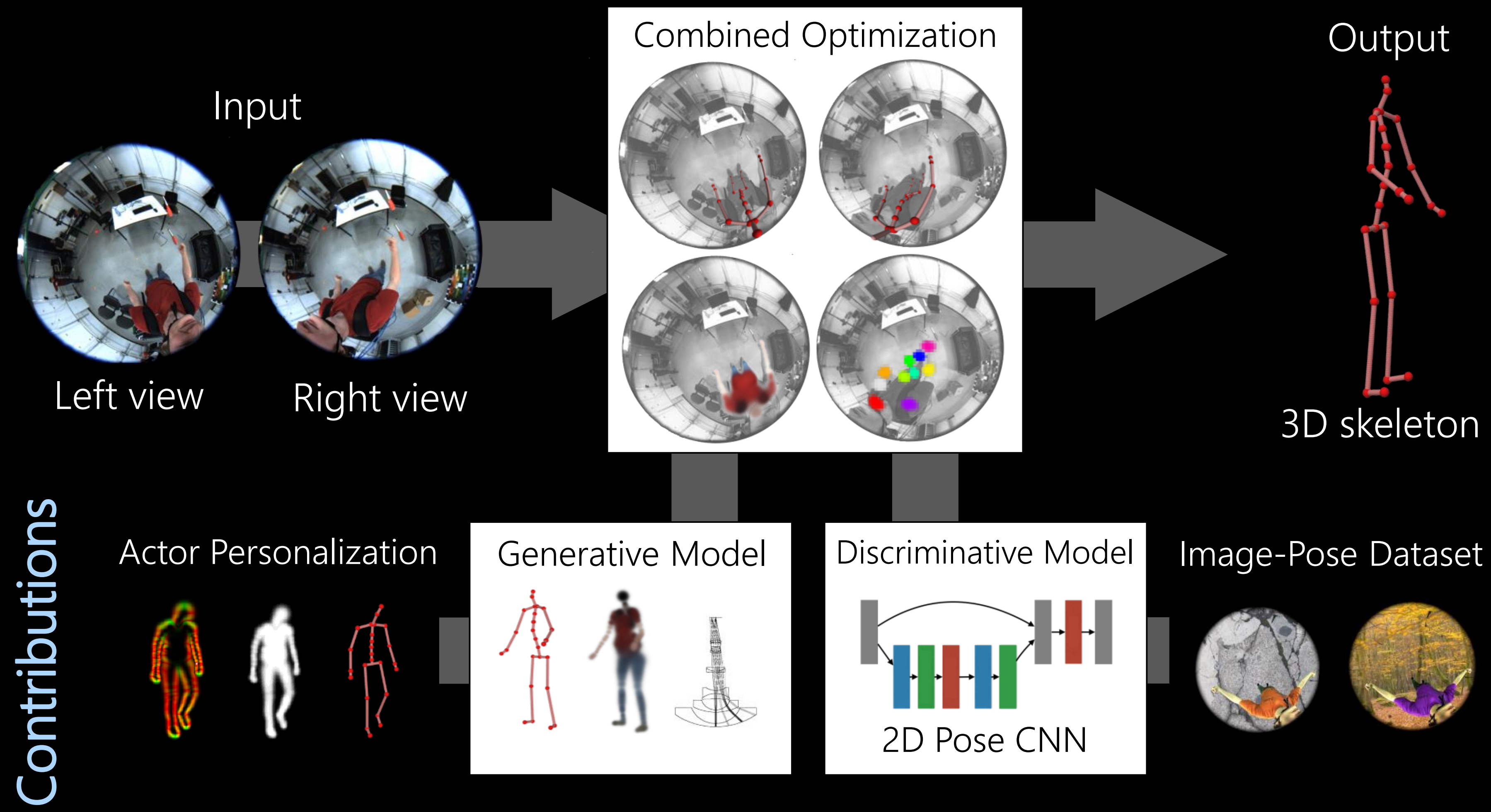
Automatic ground-truth annotation

Outside-in markerless motion capture

Projection into dynamic egocentric camera



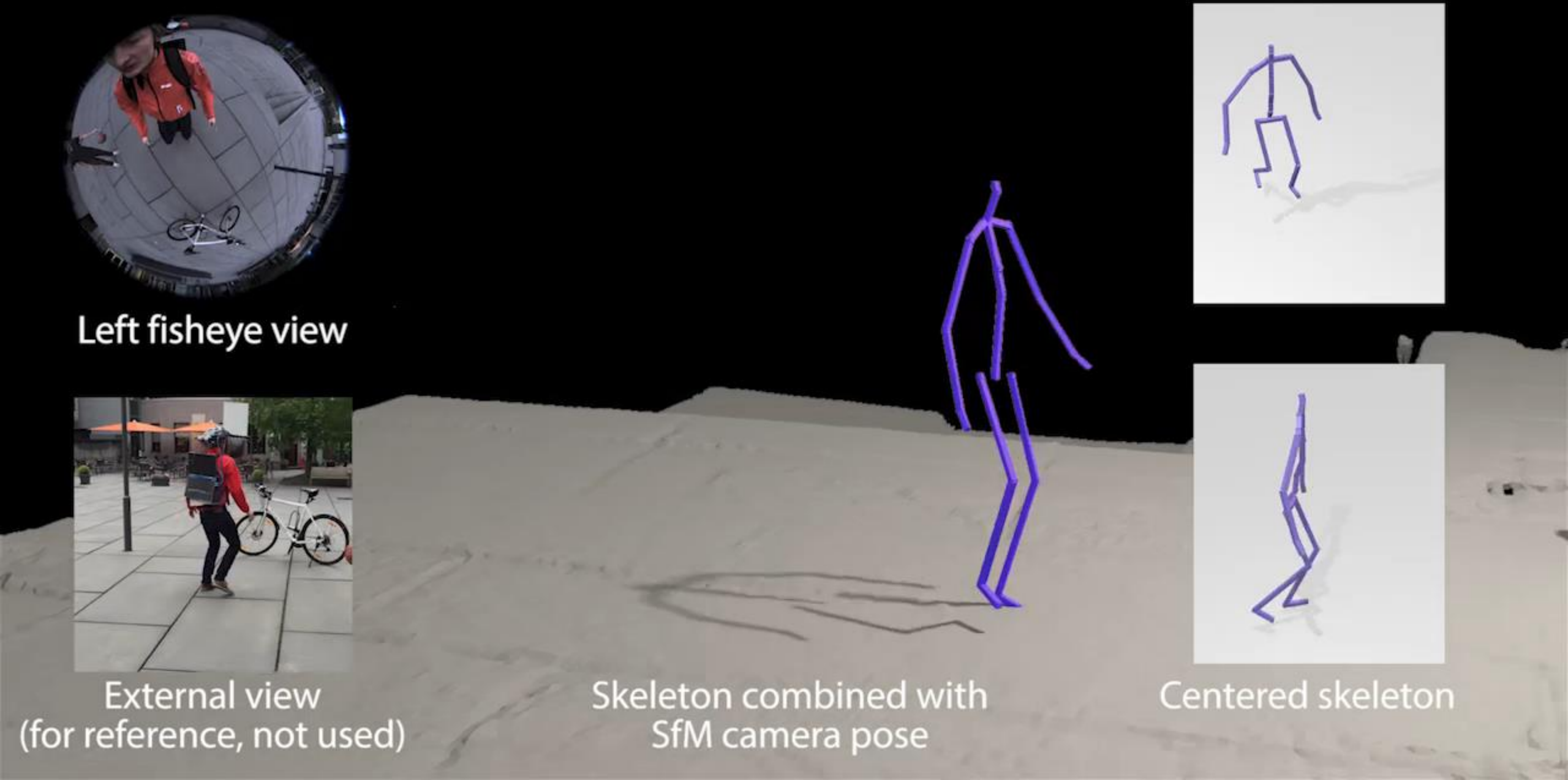
Model overview



Constrained and crowded Spaces

Two representative external views – Note the strong occlusions

Outdoor and large-scale

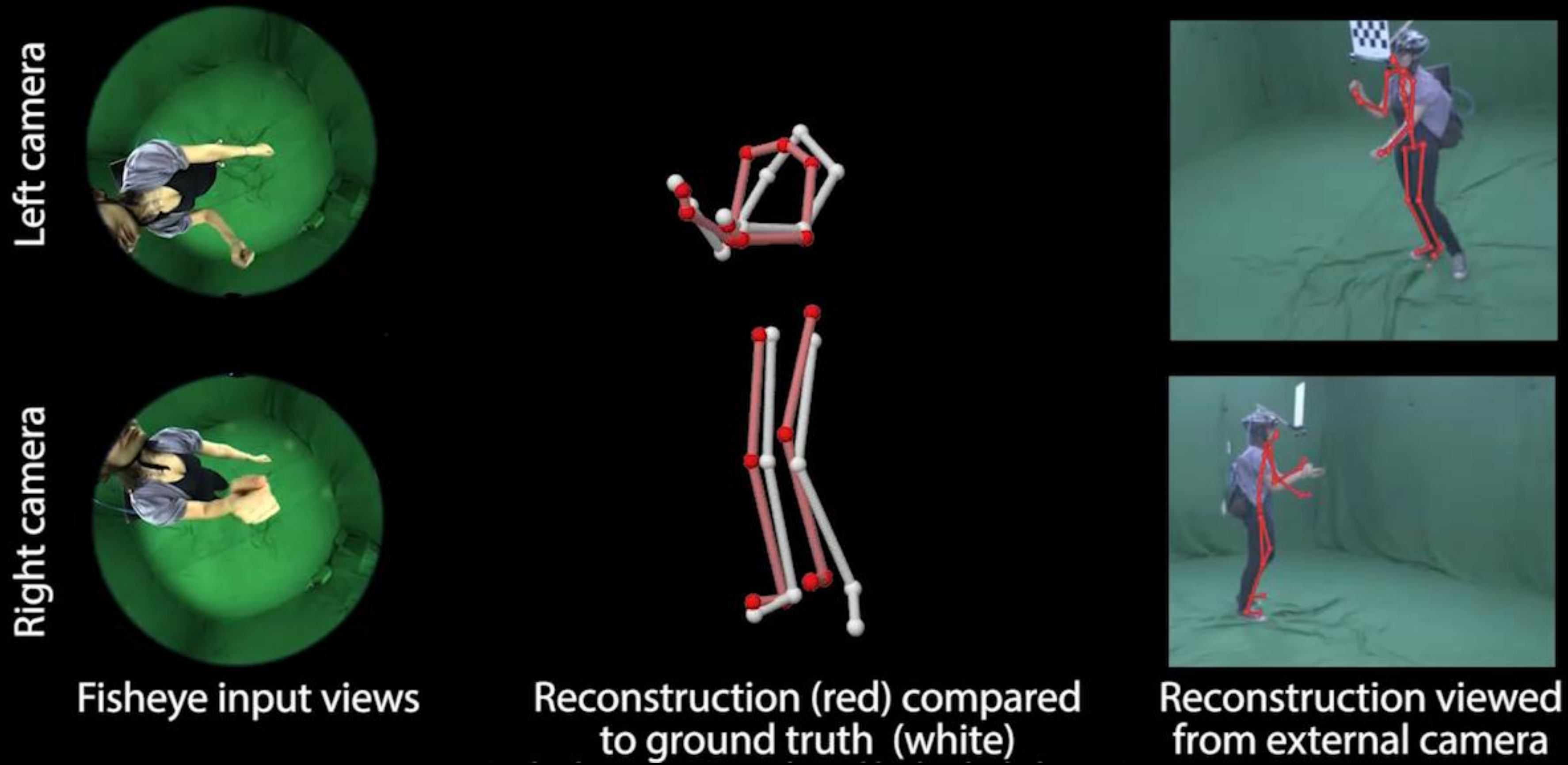


Virtual and augmented reality

Embodied virtual reality

Quantitative analysis

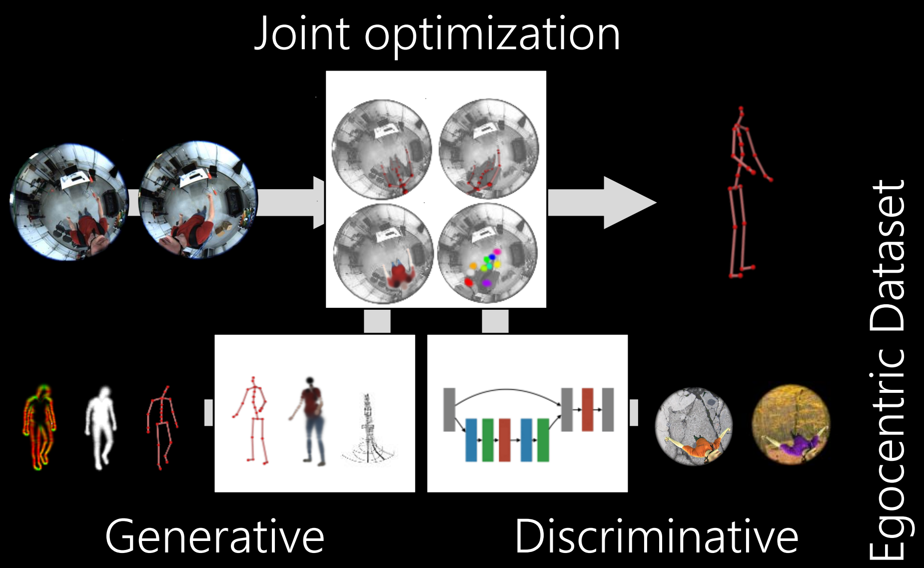
- 7 cm average Euclidean 3D error
- Temporally stable



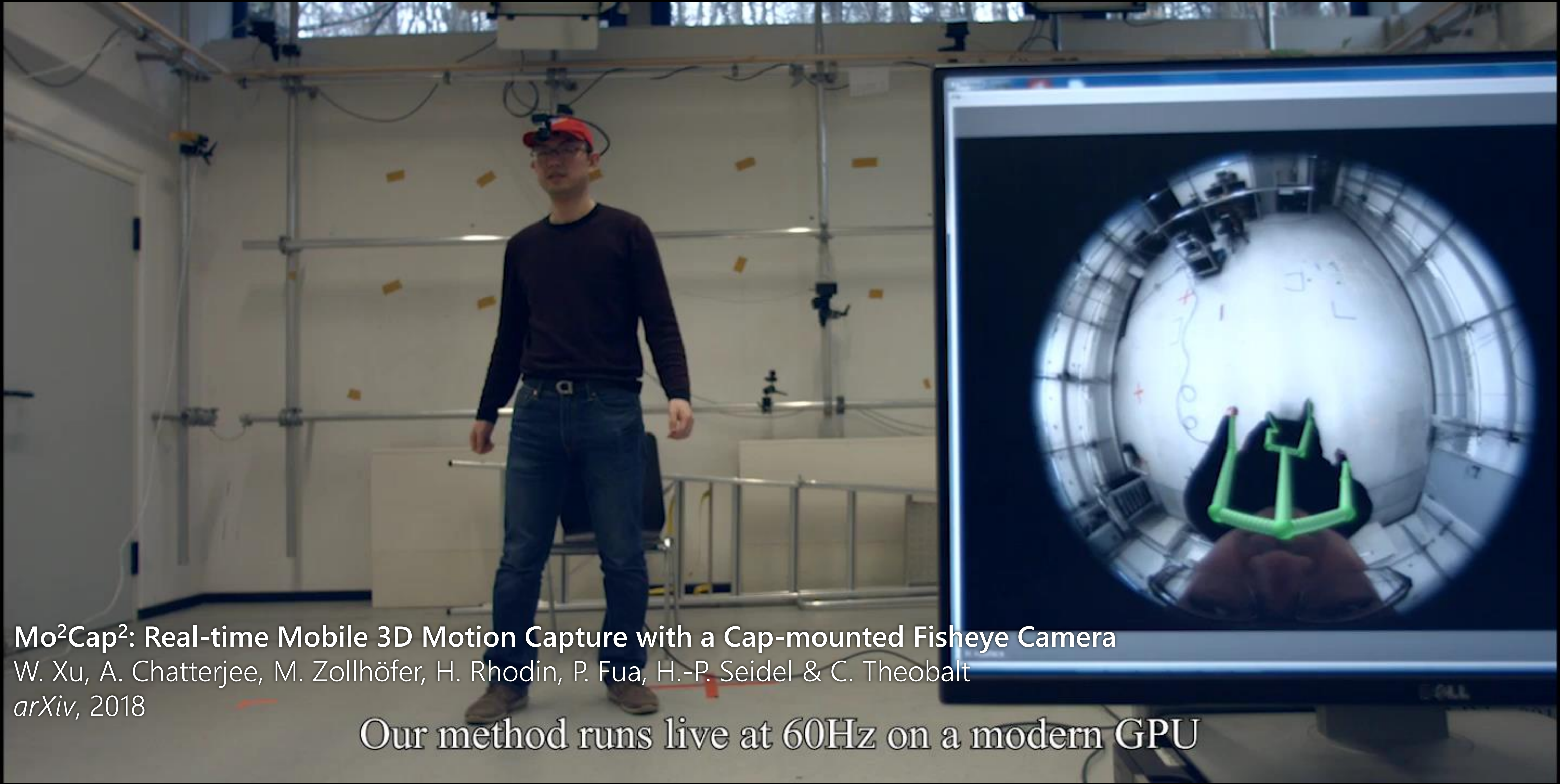
Occlusions – limitations

EgoCap summary

- Inside-in motion capture
  - full-body 3D pose
  - easy-to-setup
  - low intrusion level
  - real-time capable
  - general environments
- Future work
  - low latency (for VR)
  - alternative camera placement, monocular
  - capture hands and face



Single-camera egocentric motion capture



Mo²Cap²: Real-time Mobile 3D Motion Capture with a Cap-mounted Fisheye Camera
W. Xu, A. Chatterjee, M. Zollhöfer, H. Rhodin, P. Fua, H.-P. Seidel & C. Theobalt
arXiv, 2018

Quick recap

- Immersion & presence: motion is extremely important
  - presence breaks when visual body motion does not match physical motion
- Tacking in VR/AR: need high accuracy and update rate, low latency
  - in practice, usually best to combine IMUs with optical tracking to fix drift
- Hand input devices: controllers are tracked robustly and accurately
  - hand tracking will soon enable natural interaction with real-world objects
- Full-body motion capture: bring the entire body into VR
  - marker-based systems are fast, robust, accurate and very expensive
  - markerless systems allow live motion capture from just 1 or 2 cameras

Questions?

IEEE VR 2018 REUTLINGEN

Christian Richardt

Motion-Aware Displays

IEEE VR Tutorial on Cutting-Edge VR/AR Display Technologies

CAMERA
Centre for the Analysis of Motion, Entertainment Research and Applications

UNIVERSITY OF BATH