Interacting with Cameras

Hrvoje Benko
Microsoft Research

UW CSE 510 – February 9, 2016

Papers assigned

Pierre Wellner. 1993. Interacting with paper on the DigitalDesk. Communications of the ACM 36, 7 (July 1993), 87-96. DOI=10.1145/159544.159630 http://doi.acm.org/10.1145/159544.159630

Pick one:

Hrvoje Benko, Ricardo Jota, and Andrew Wilson. 2012. MirageTable: Freehand Interaction on a Projected Augmented Reality Tabletop. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '12). ACM, New York, NY, USA, 199-208. DOI=http://dx.doi.org/10.1145/2207676.2207704

Andrew D. Wilson and Hrvoje Benko. 2014. CrossMotion: Fusing Device and Image Motion for User Identification, Tracking and Device Association. In *Proceedings of the 16th International Conference on Multimodal Interaction* (ICMI '14). ACM, New York, NY, USA, 216-223. DOI=http://dx.doi.org/10.1145/2663204.2663270

Class Outline

- Motivation
- Basic camera processing
- Applications:
  - Cameras as rapid prototyping tools
  - Touch tracking
  - 3D gestures + mid-air interactions
  - Device tracking/sensor fusion
- Challenges with Designing Sensing Systems

WHY SENSE WITH CAMERAS?

Cameras enable novel interactive experiences

Visual appearance

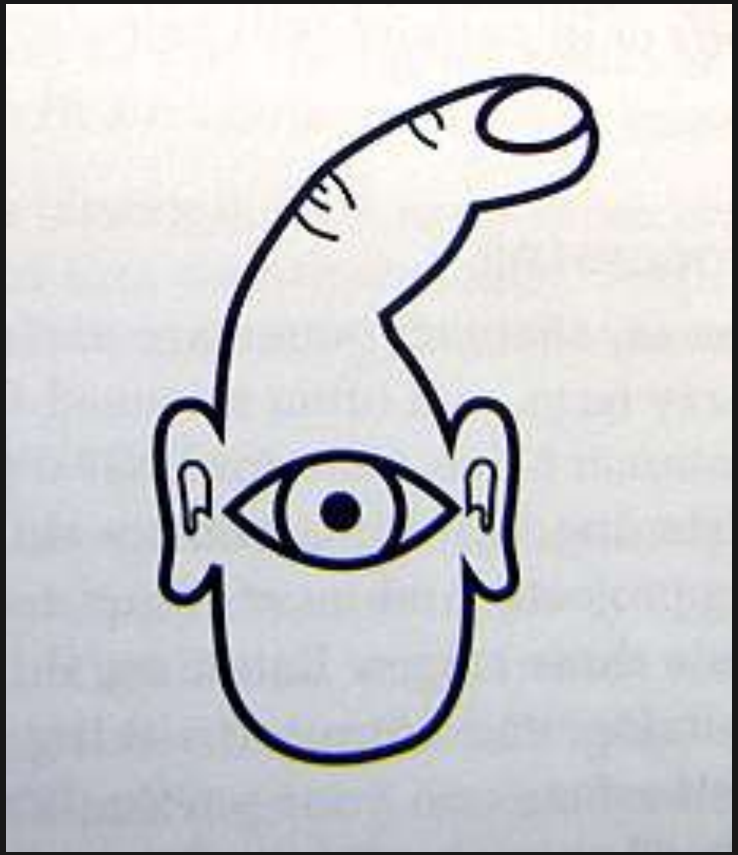
Call of Duty Modern Warfare 2

Visual appearance

Interactivity

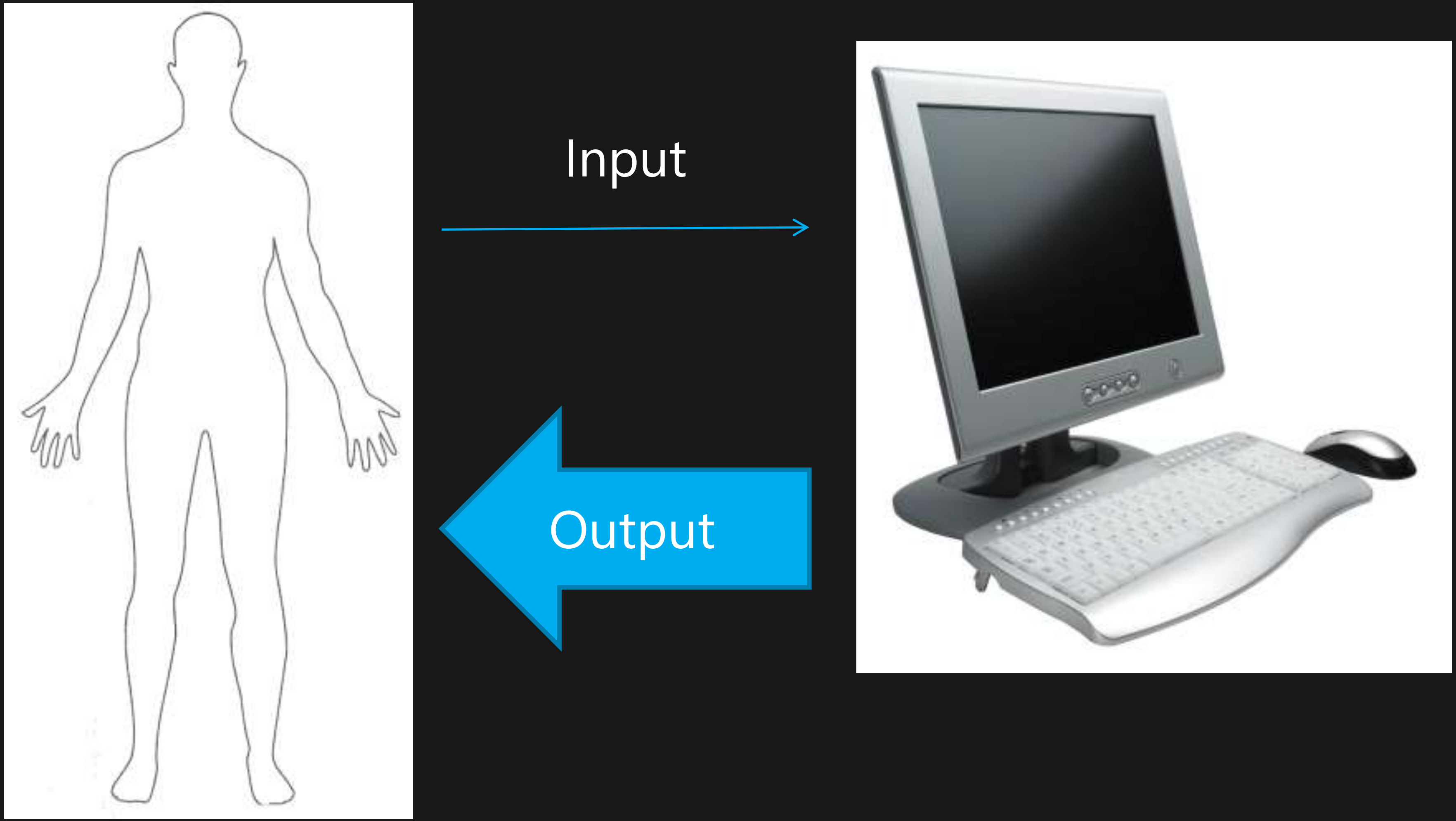
How we see the computer!

How the computer sees us!

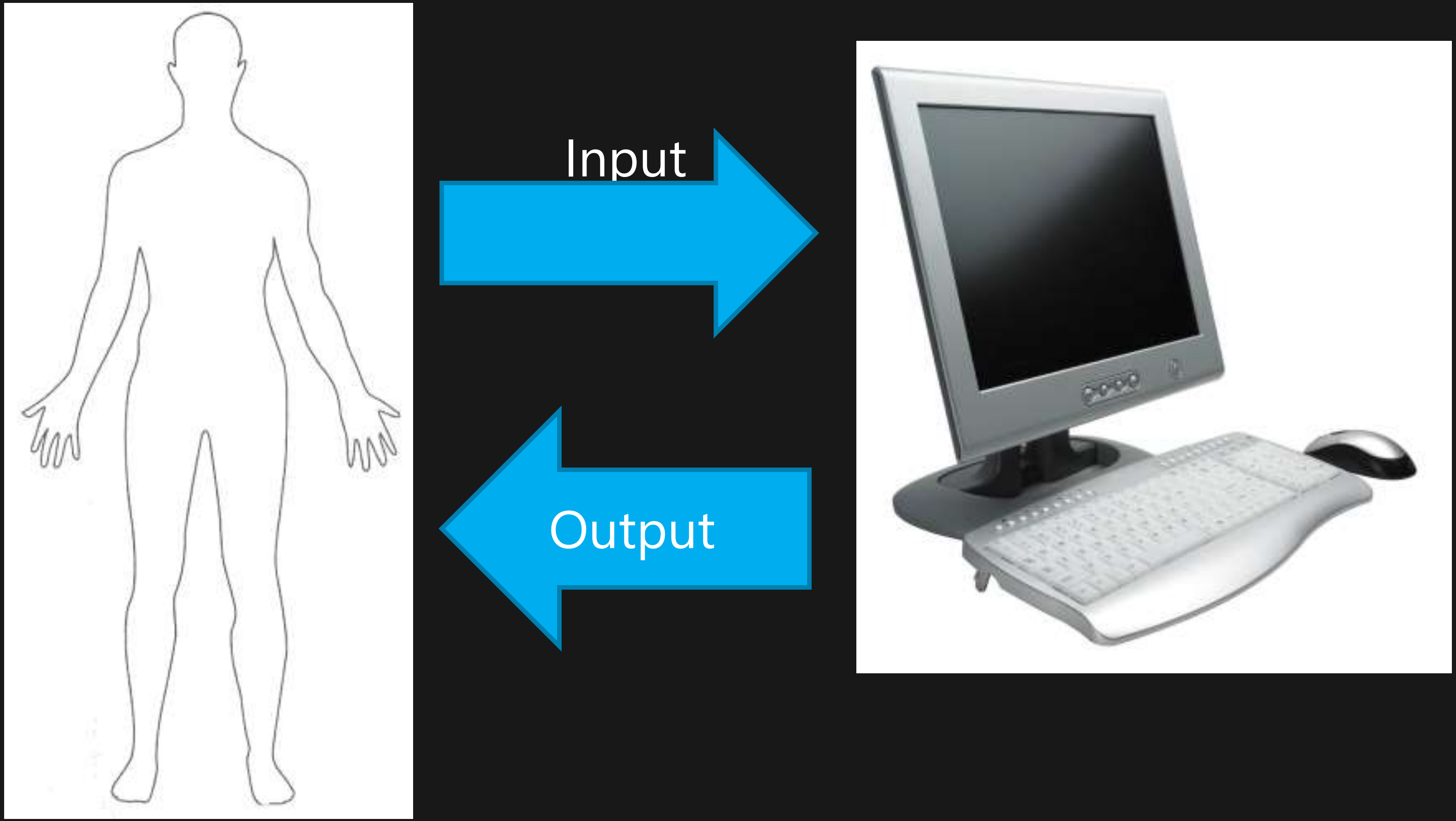


Tom Igoe and Dan O’Sullivan - *Physical Computing*.

Human Computer Interaction



Human Computer Interaction

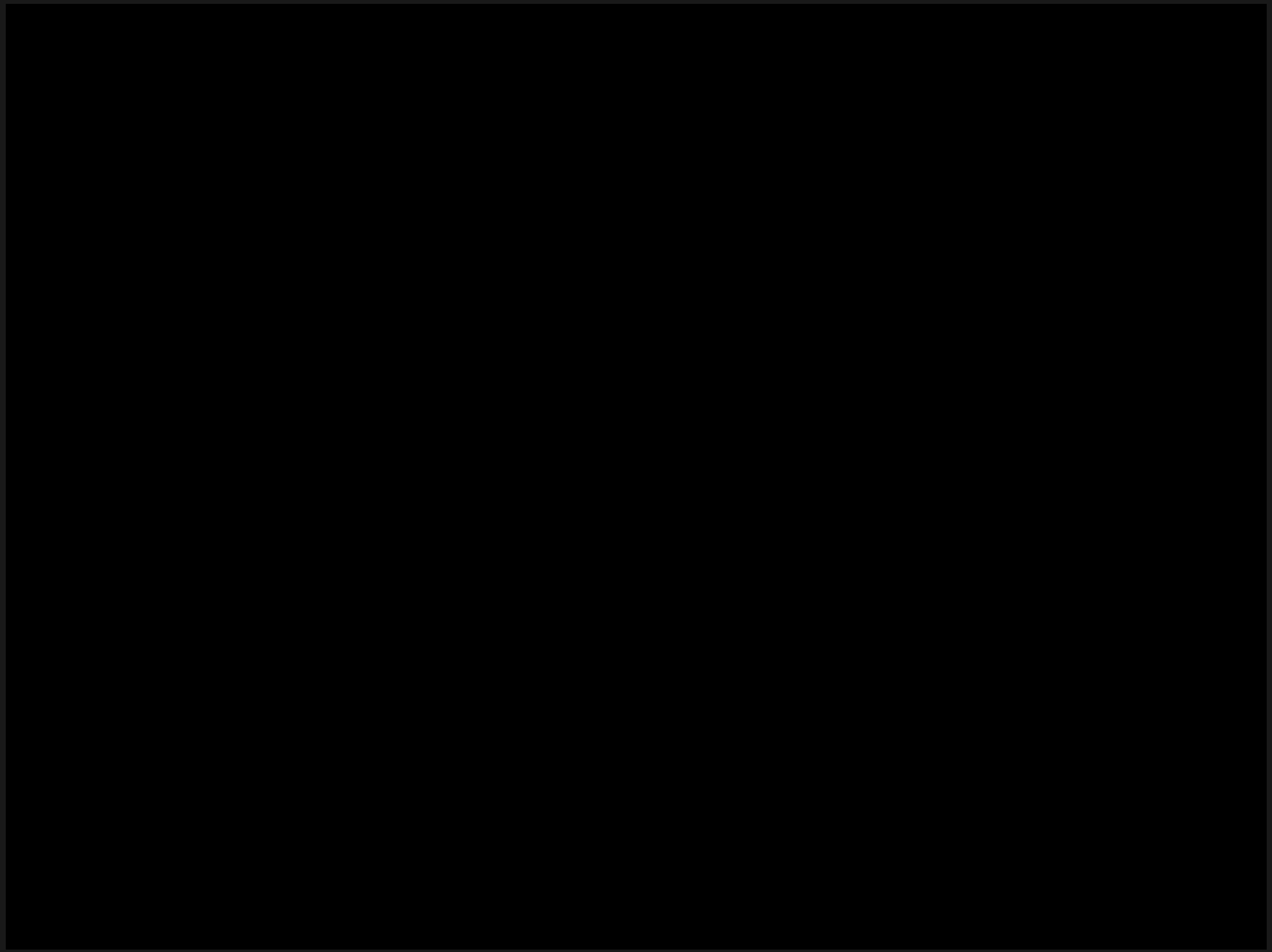


Myron Krueger's Videoplace

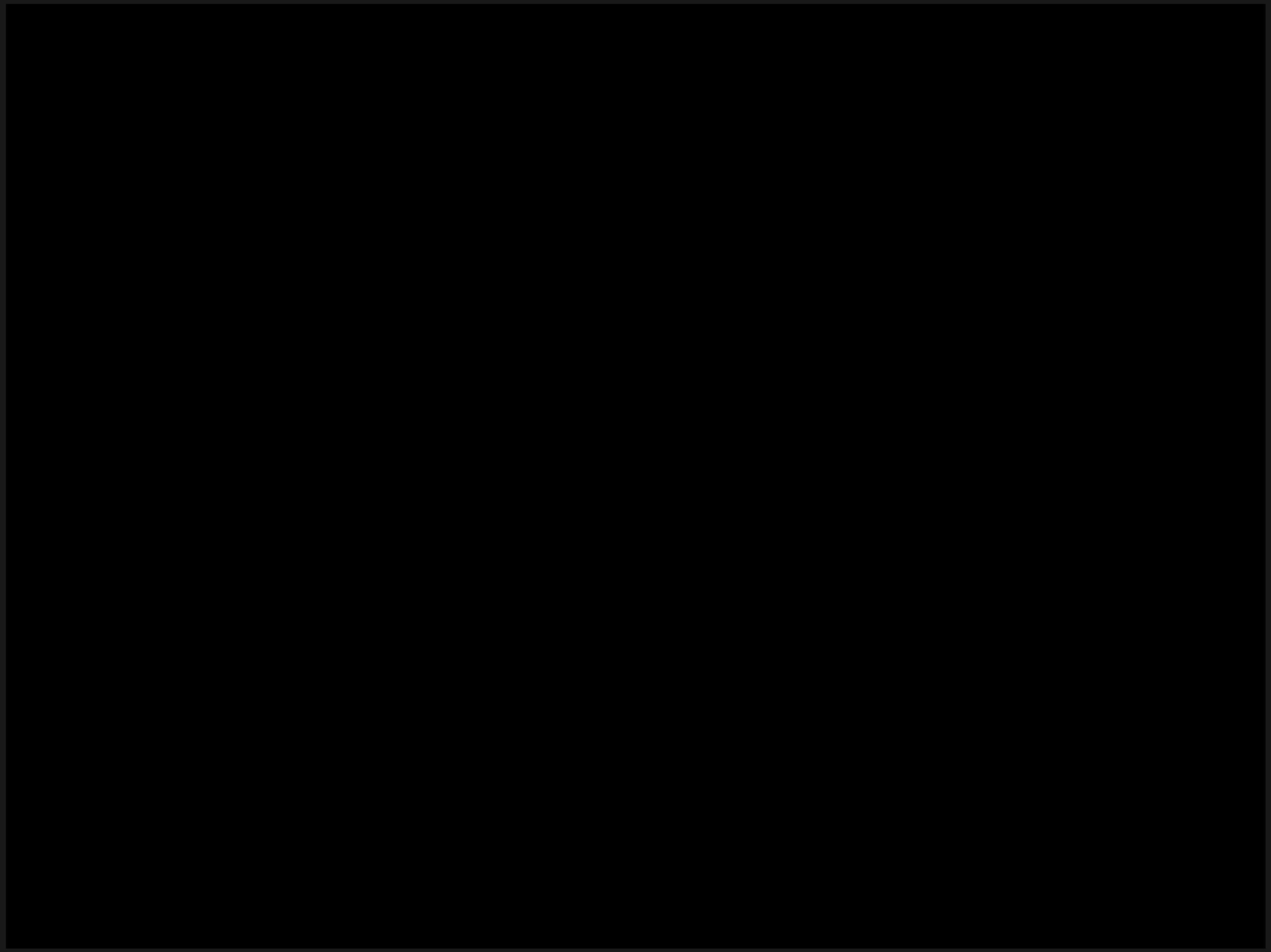
MicroMotoCross

A. Wilson, ACM ITS 2007

Beach Ball



DigitalDesk



Pierre Wellner. 1993. Interacting with paper on the DigitalDesk. Communications of the ACM 36, 7 (July 1993), 87-96. DOI=10.1145/159544.159630 http://doi.acm.org/10.1145/159544.159630

Discussion

How much of this vision is realized today?

What about occlusions, accidental activation?

Do we need this? Are we close to “paperless office”?

What about AR/VR?

IMAGE PROCESSING PRIMER

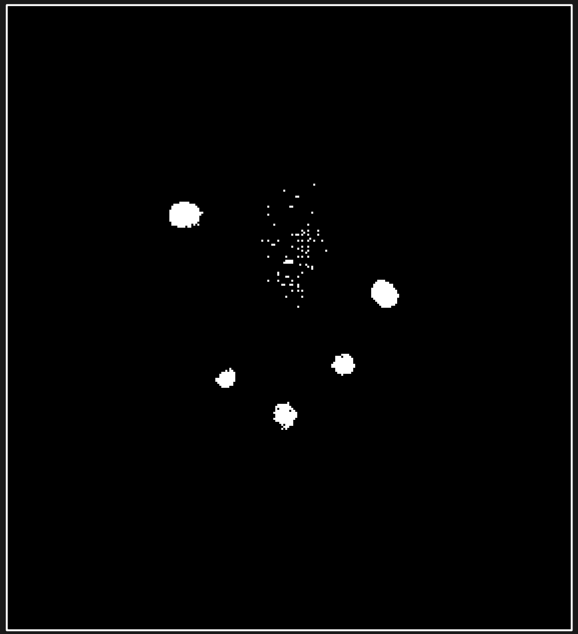
Typical Pipeline

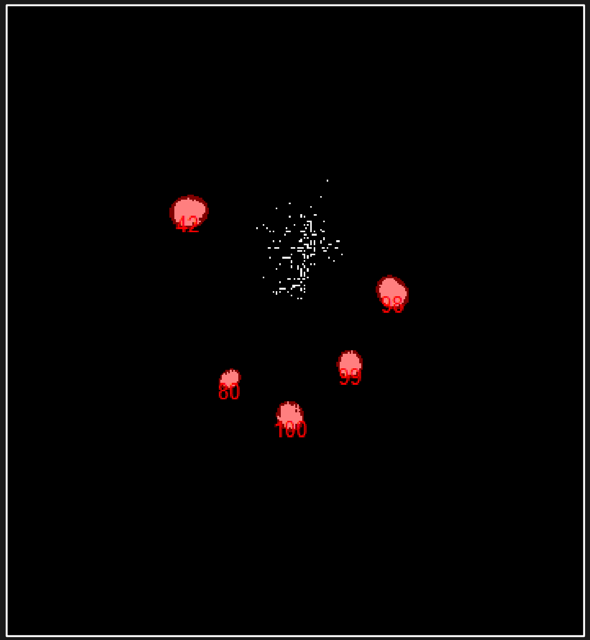
1. Segment foreground from background
2. Cluster foreground pixels into “blobs”
3. Track blobs over time

Image Segmentation

Thresholding based on intensity (or color)

Blob tracking





Surface Computing

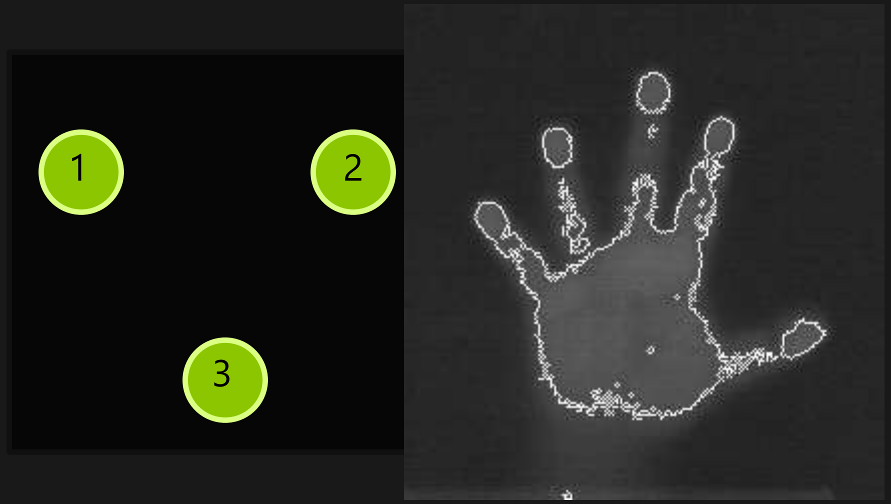
MS Surface

PerceptivePixel

Cursors considered harmful

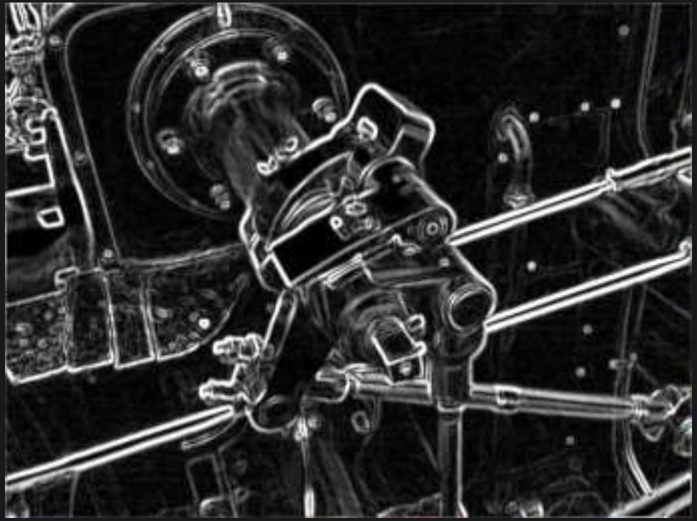
Pulling out discrete contact points is an *ill-posed* problem

Leads to all sorts of mayhem!

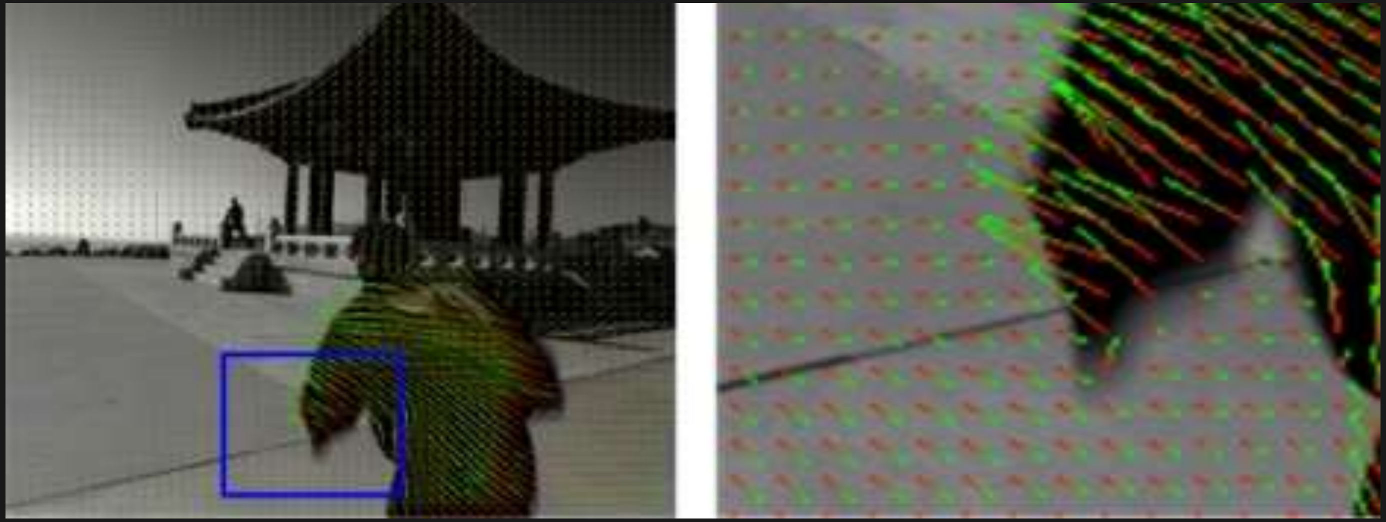


Other things to track

Edges

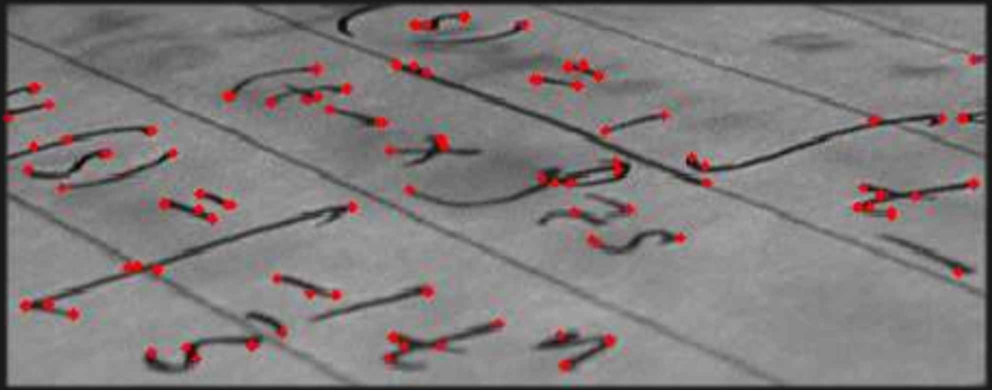


Motion



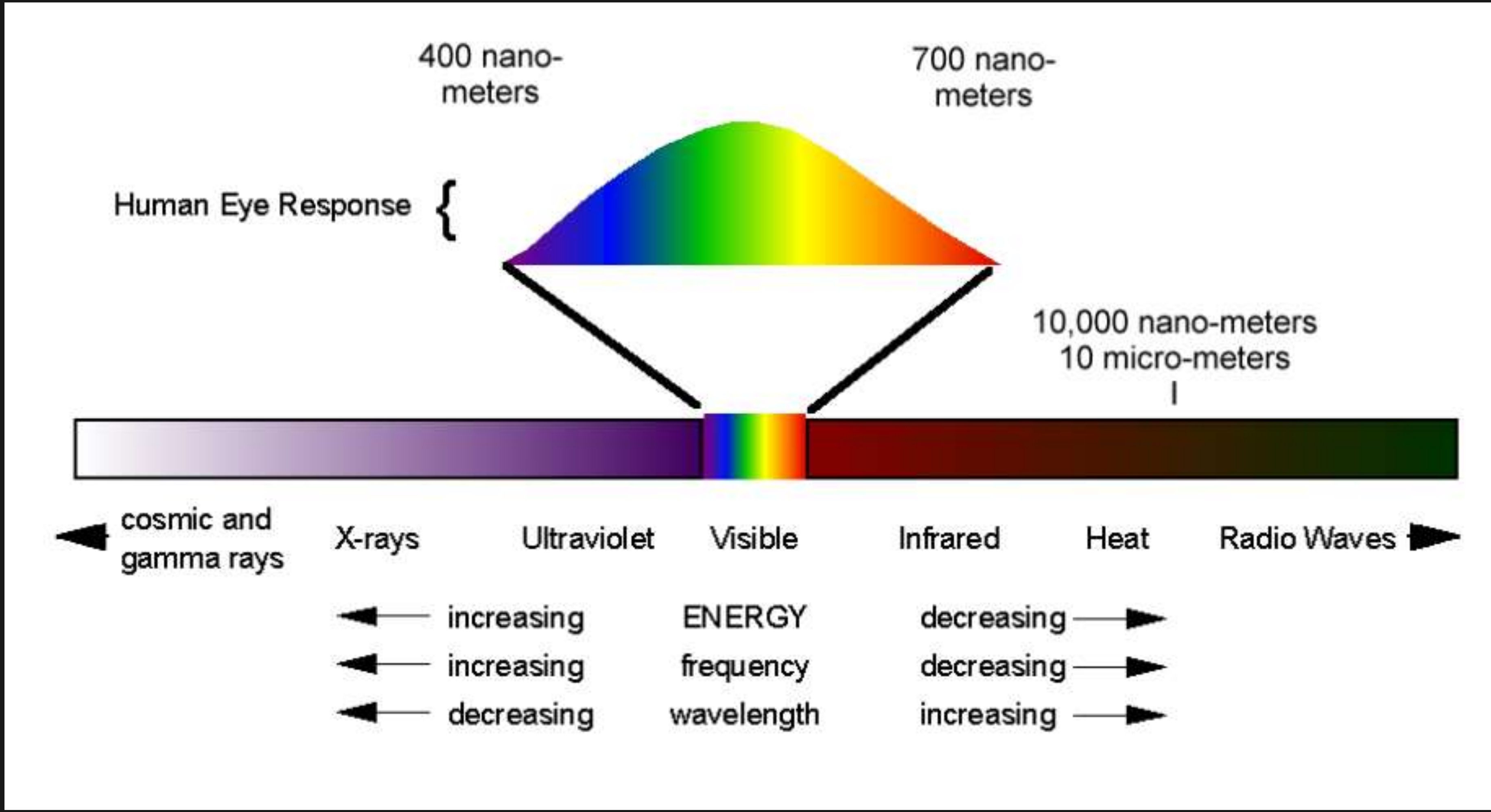
Other things to track

Features
Corners, lines,
SIFT, SURF, etc.

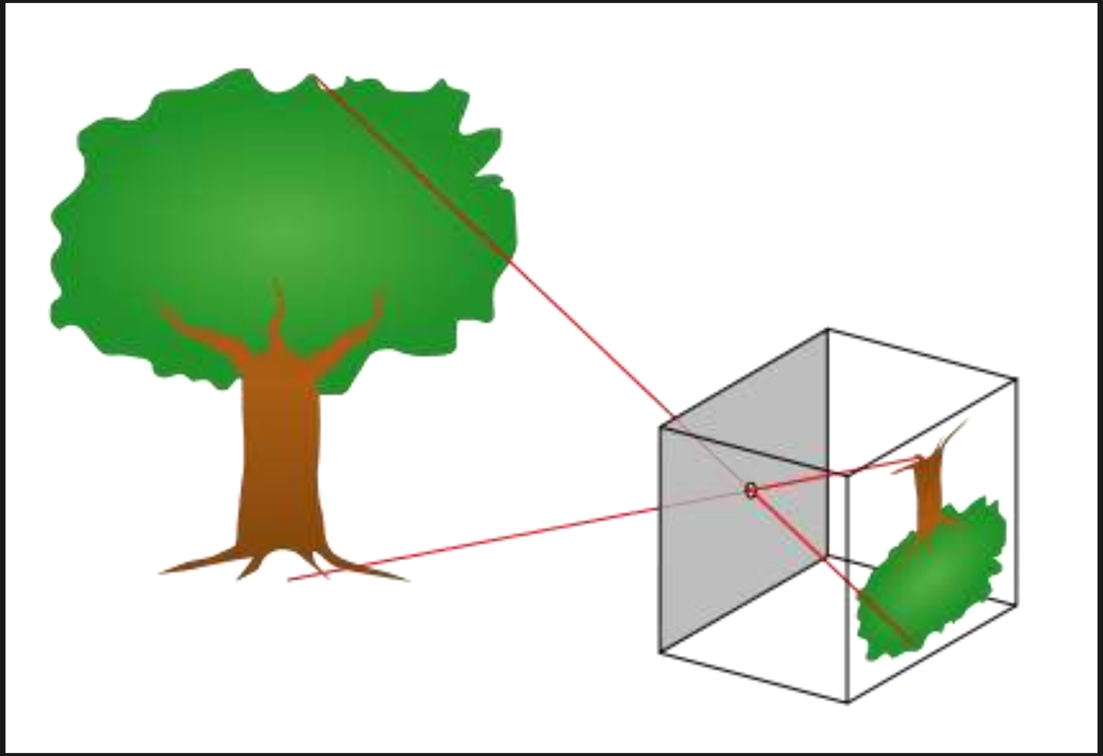


Trick of the trade: Spectrum separation

Move sensing into infrared!

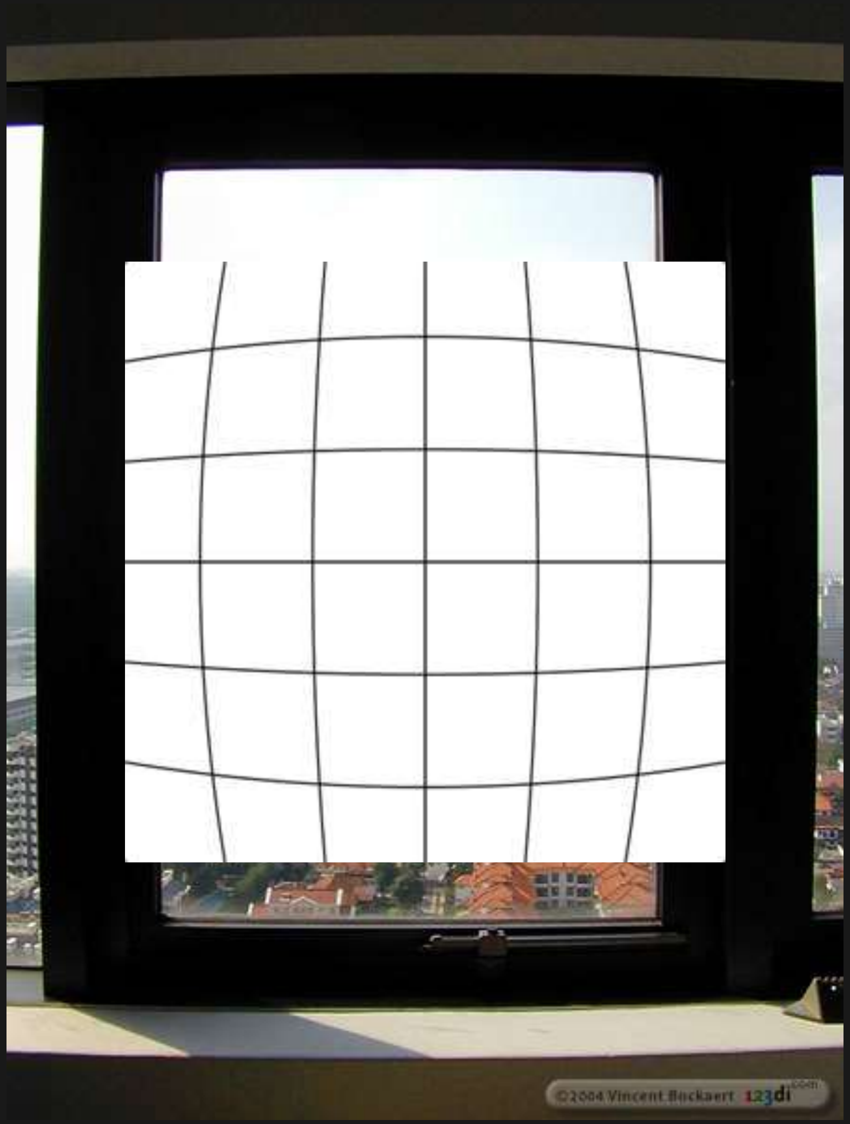


Ideal Pinhole Camera Model



Real cameras

Have lenses
Distortions
Focal lengths
Principle points



Calibration

Requires defining:

- Intrinsic parameters (focal length, image format, principal point)
- Lens distortion (usually non-linear)
- Extrinsic parameters (R, T)

Most projects start by using:

Z. Zhang, "A flexible new technique for camera calibration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

http://research.microsoft.com/en-us/um/people/zhang/Papers/TR98-71.pdf

NON-STANDARD “CAMERAS”

Touch Sensors as Cameras

Rekimoto, SmartSkin, CHI 2002

Projectors as Inverted Cameras

Wilson, PlayAnywhere, UIST 2005

Fast Framerate

ROI tricks on CMOS cameras

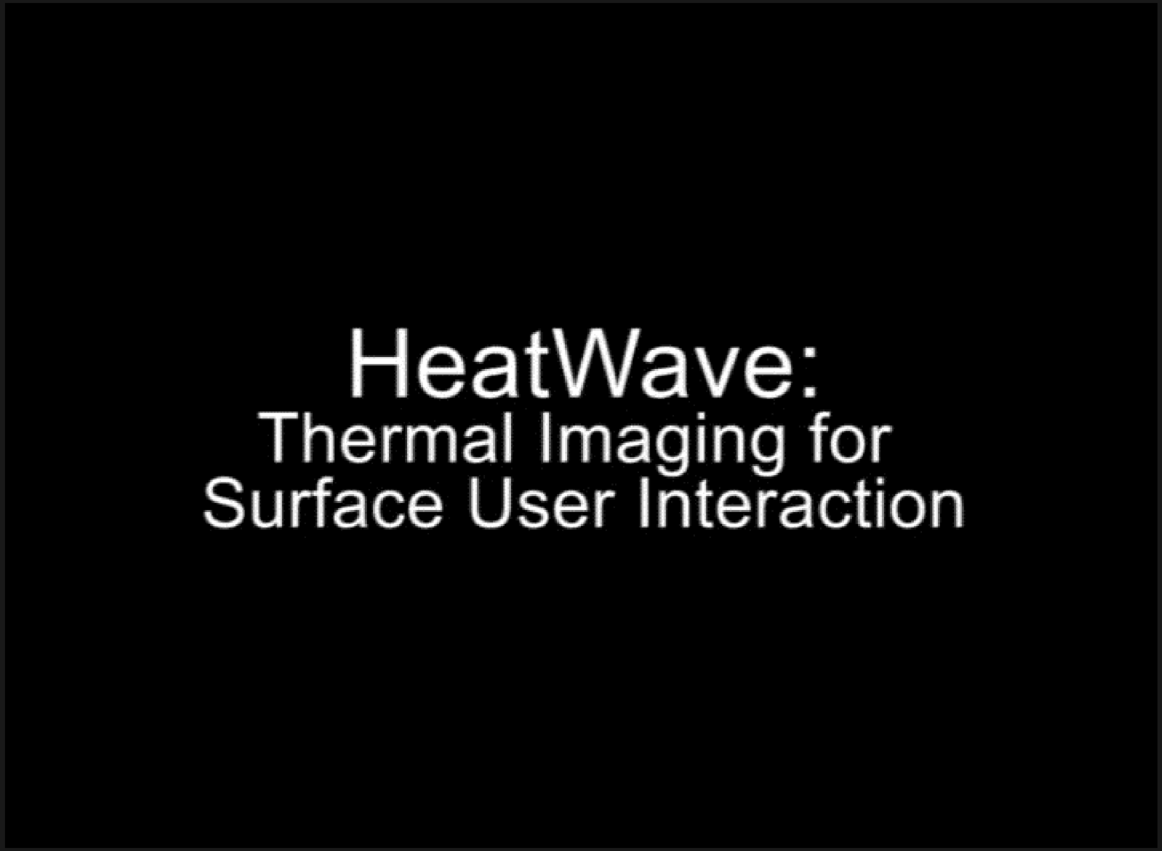
Line cameras (PhaseSpace)

Near vs. Far Infrared

Requires light source
Reflected radiation

Thermal imaging
Emissive radiation

HeatWave



Larson, E., Cohn, G., Gupta, S., Ren, X., Harrison, B., Fox, D., Patel, S.N. **HeatWave: Thermal Imaging for Surface User Interaction**. In the *Proceedings of CHI 2011* (May 7-12, Vancouver, Canada), ACM, New York, 2011, pp. 2565-2574.

Depth sensing cameras

Color + depth per pixel: RGBZ

Can compute world coordinates of every point in the image directly.



Depth Camera Flavors

- Stereo
- Structured light
- Time of flight

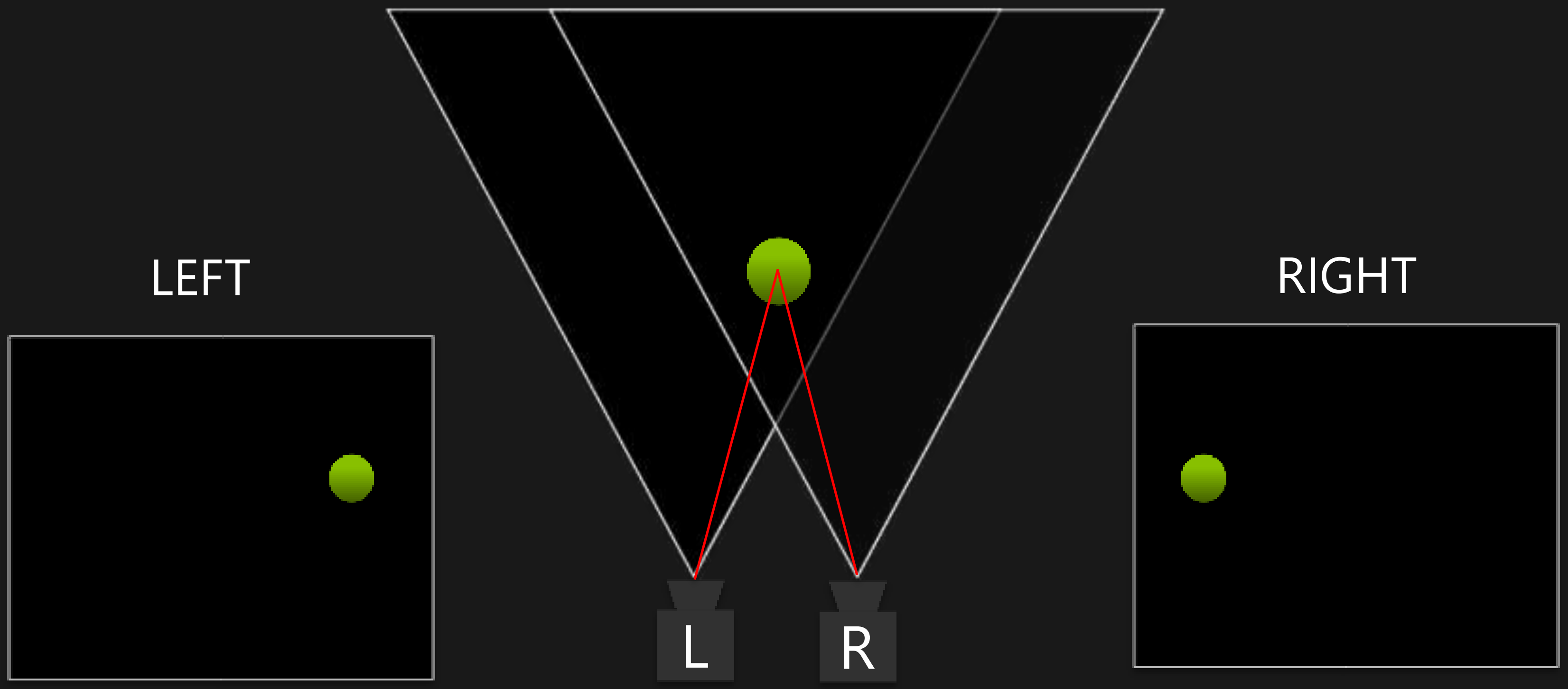
Correlation-based stereo cameras

Binocular disparity

ZED Stereo Camera www.stereolabs.com

Point Grey Research www.ptgrey.com

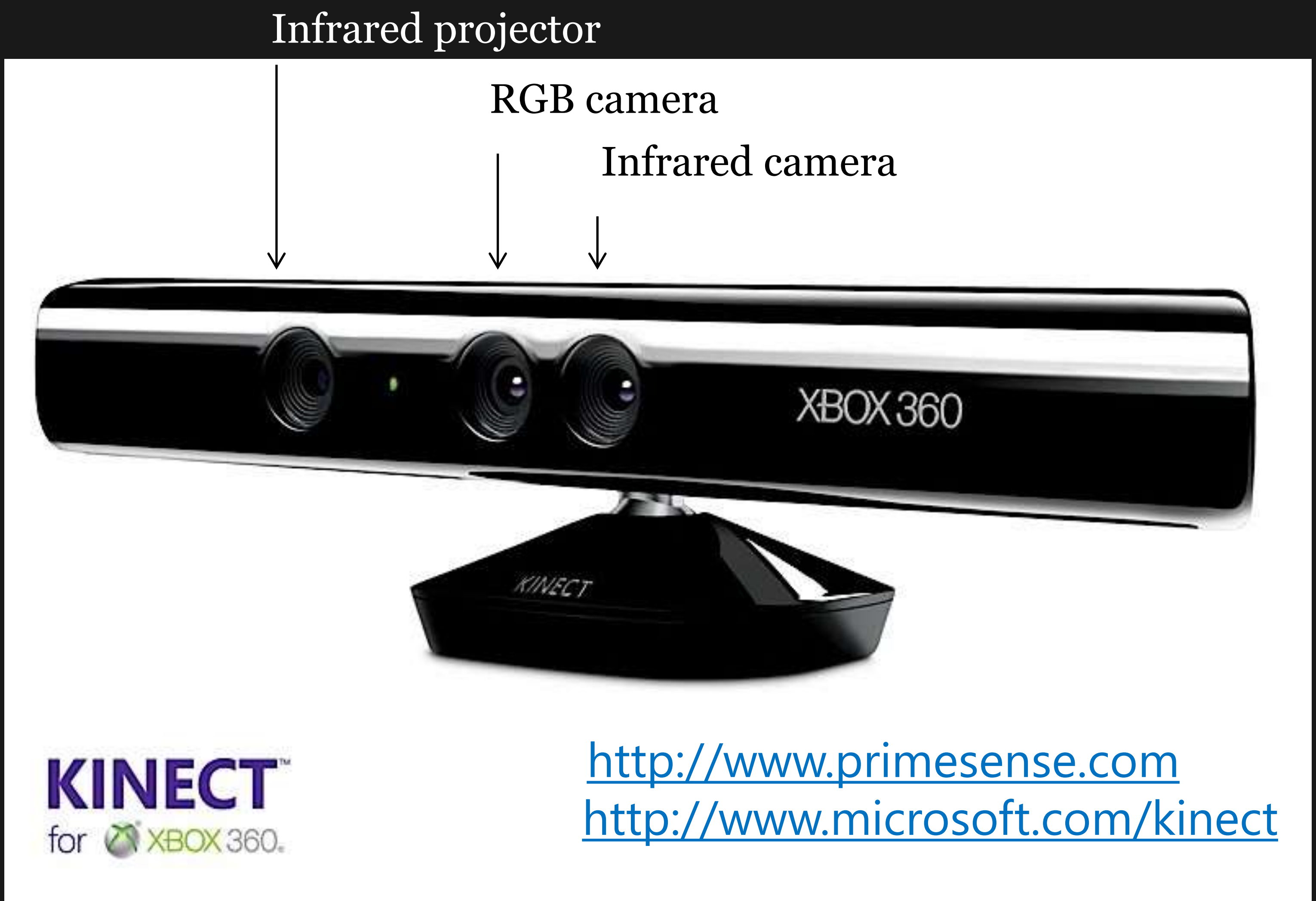
Correlation-based stereo



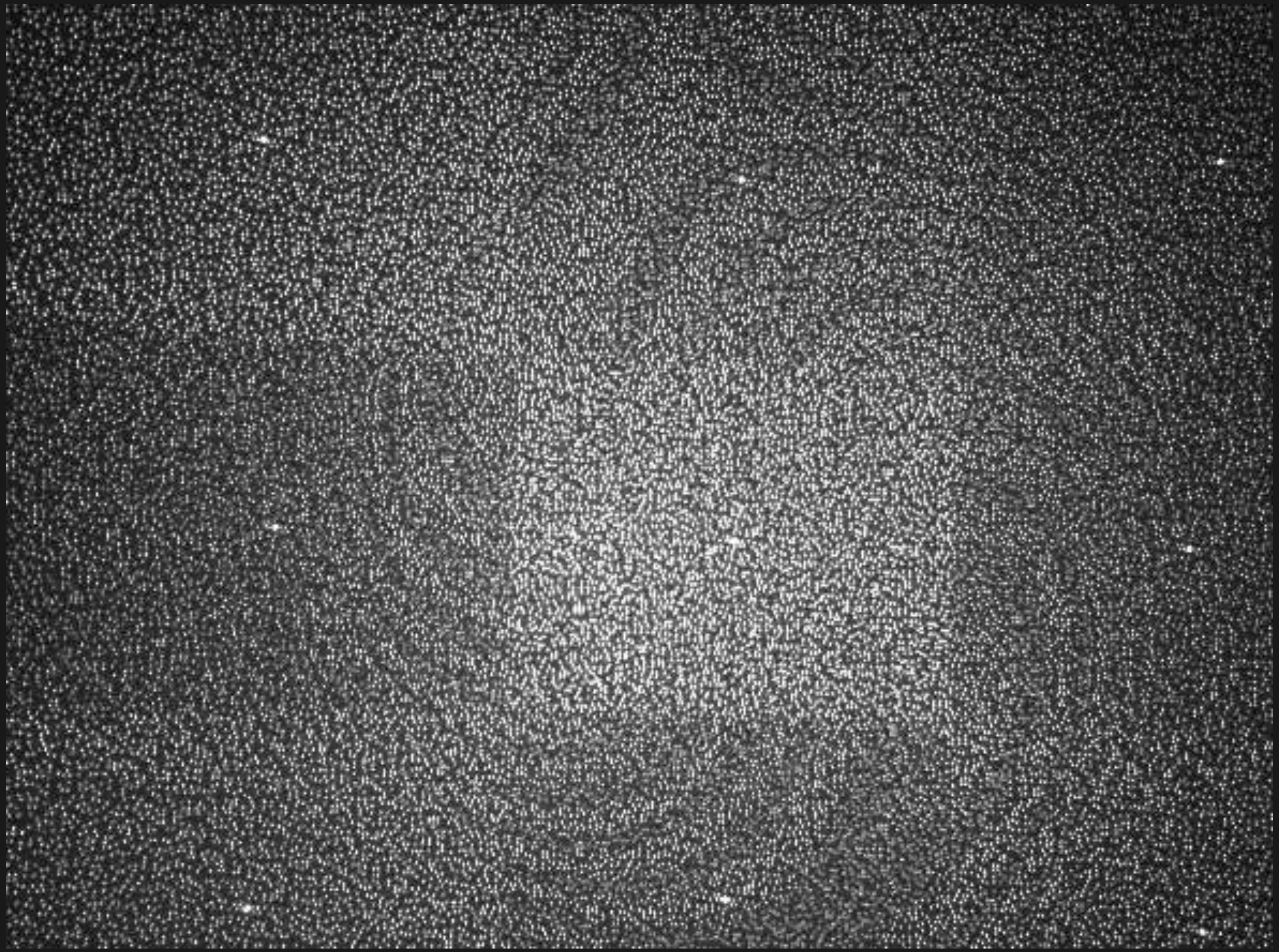
Stereo drawbacks

- Requires good texture to perform matching
- Computationally intensive
- Fine calibration required
- Occlusion boundaries
- Naïve algorithm very noisy

Structured light depth cameras

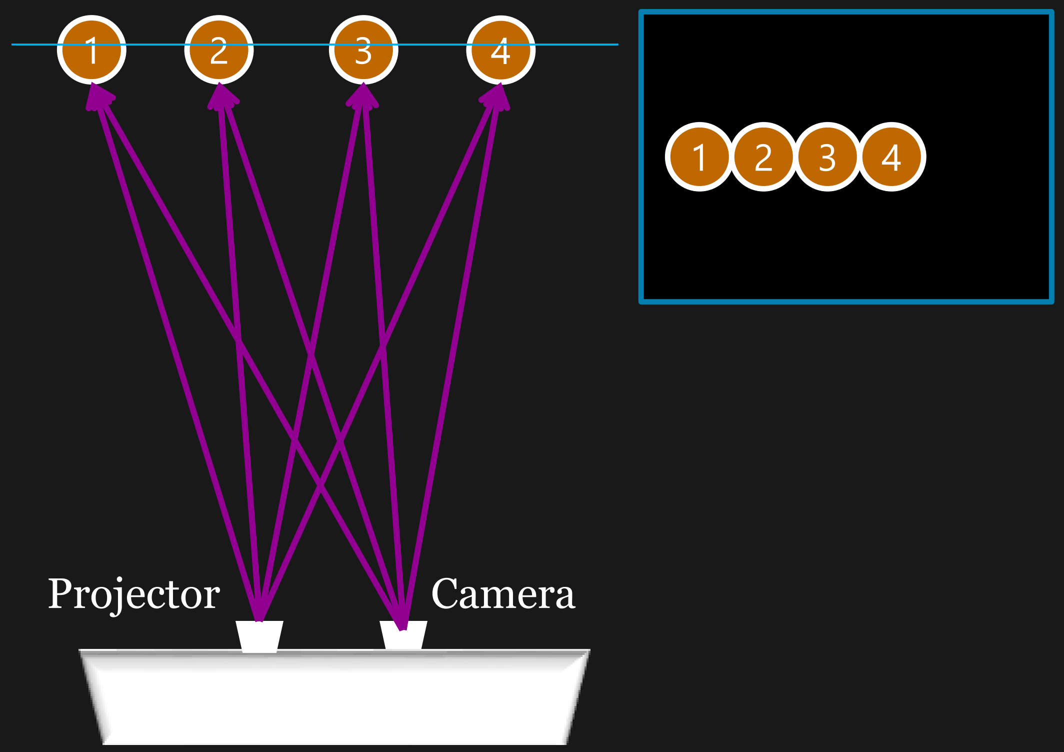


Known Projected IR pattern



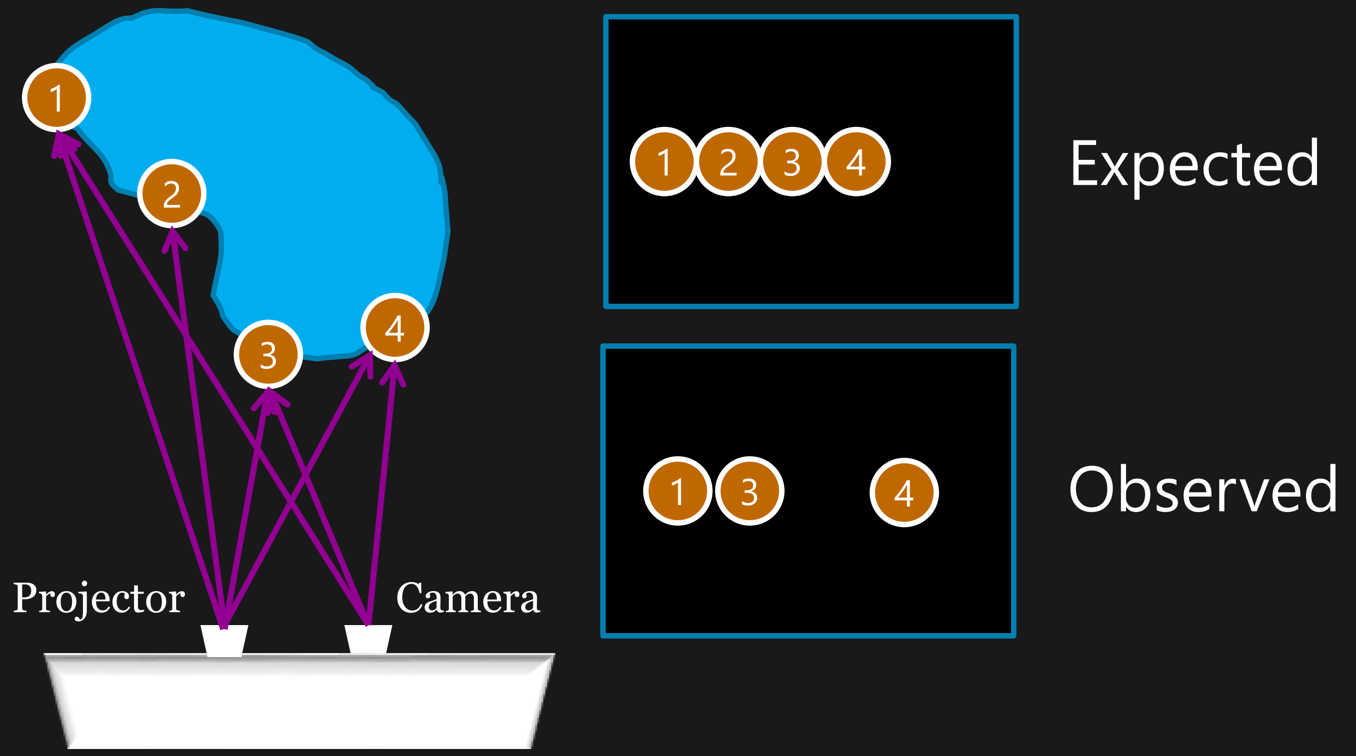
Object = disturbance

Depth by structured light coding



- Expect a certain pattern at a given point

Depth by structured light coding



- Expect a certain pattern at a given point
- Find how far this pattern has shifted
- Relate this shift to depth (triangulate)

Kinect v1

Fast

Computationally inexpensive

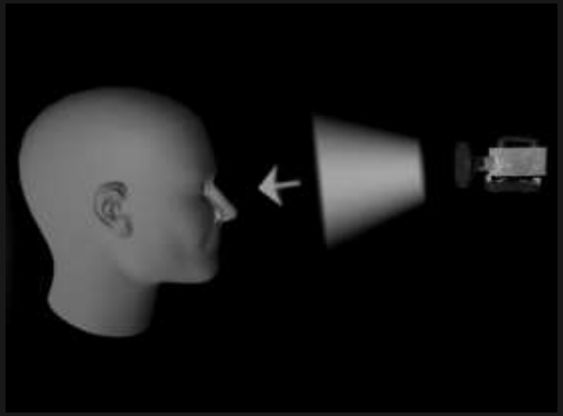
The object needs to be big enough to contain enough pattern to encode itself

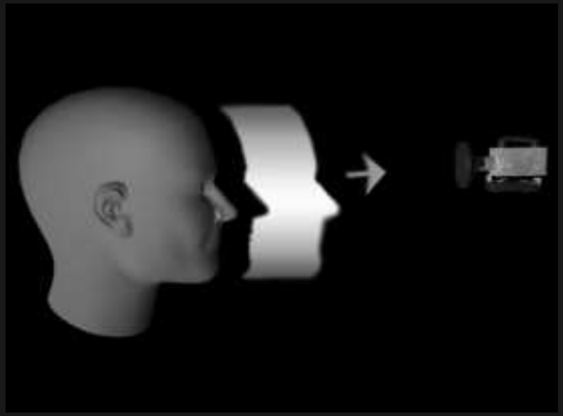
- Edges noisy, surfaces smooth

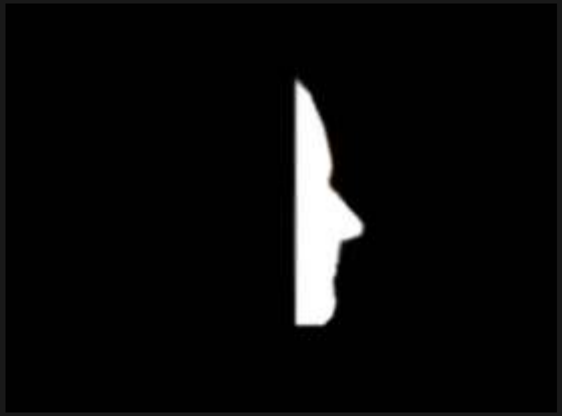
Time of Flight Depth Cameras

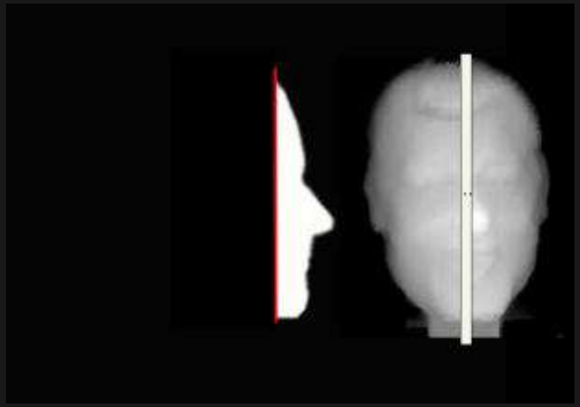
Kinect v2
3DV, Canesta (no-longer public)
SoftKinectic http://www.softkinetic.com
PMD Technologies http://www.PMDTec.com
Mesa Technologies http://www.mesa-imaging.ch

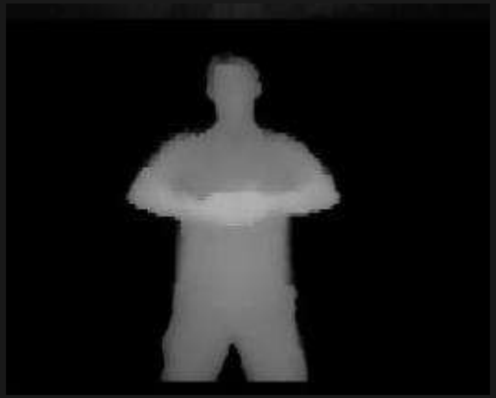
Time of flight measurement











Why sense with depth cameras?

Easier segmentation

Easier understanding of physical objects in space.

Real world units.

Skeletal tracking (from limited viewpoints).

Applications

CAMERAS ENABLE RAPID PROTOTYPING OF INTERACTIVE DEVICES

Sauron

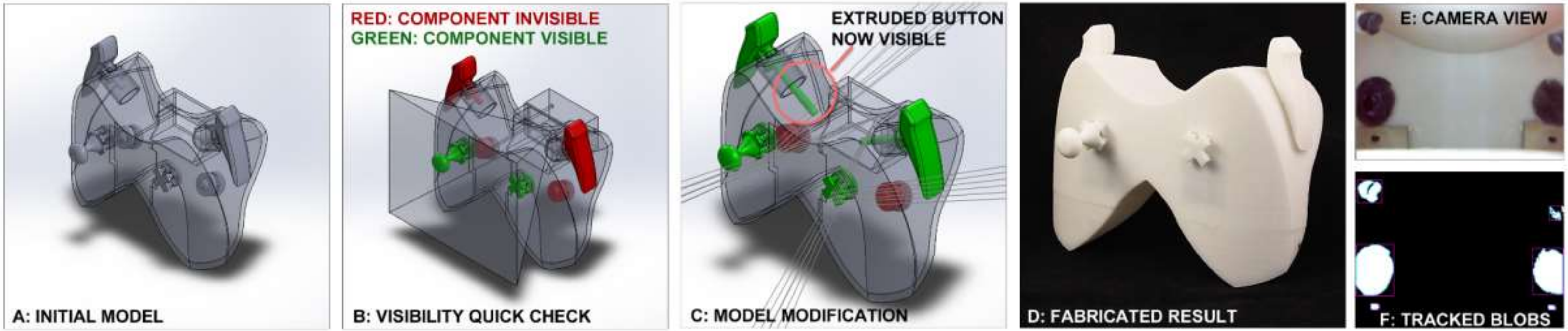


Figure 2. When designing with Sauron, a designer begins with his model (A), then inserts a virtual camera and runs quick check for visibility (B). A full model modification pass (C) performs extrusions and suggest mirror placement to bring invisible controls into the camera's view. He fabricates his design (D), then colors the inside and inserts the camera and mirrors (E). The computer vision software tracks the motion of components (F) and forwards events on to control software, such as a game.

Valkyrie Savage, Colin Chang, and Björn Hartmann. 2013. Sauron: embedded single-camera sensing of printed physical user interfaces. In *Proceedings of the 26th annual ACM symposium on User interface software and technology* (UIST '13). ACM, New York, NY, USA, 447-456. DOI=http://dx.doi.org/10.1145/2501988.2501992

Sauron video

INTERACTIONS IN MID AIR

Minority Report

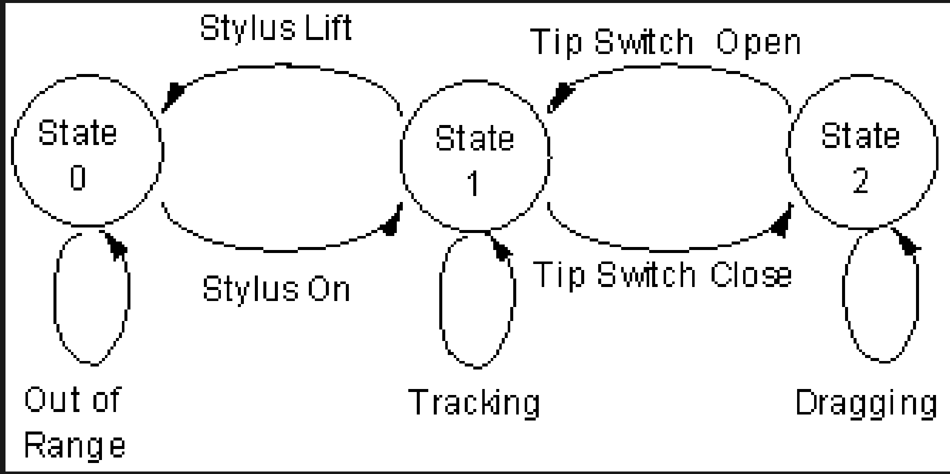
2002

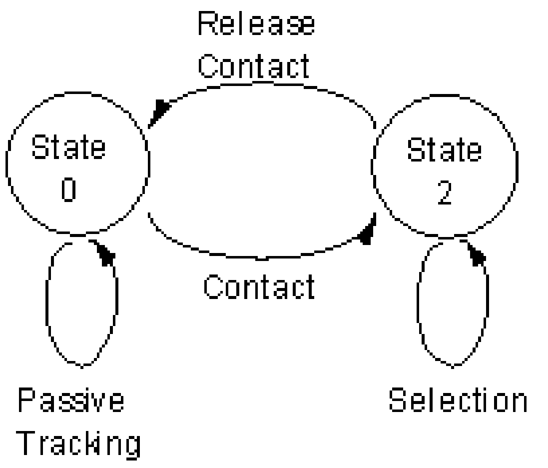
Challenges

Fatigue
Precision
Feedback
Selection

Three vs. Two states model

- How to “click”?
- How to “clutch”?





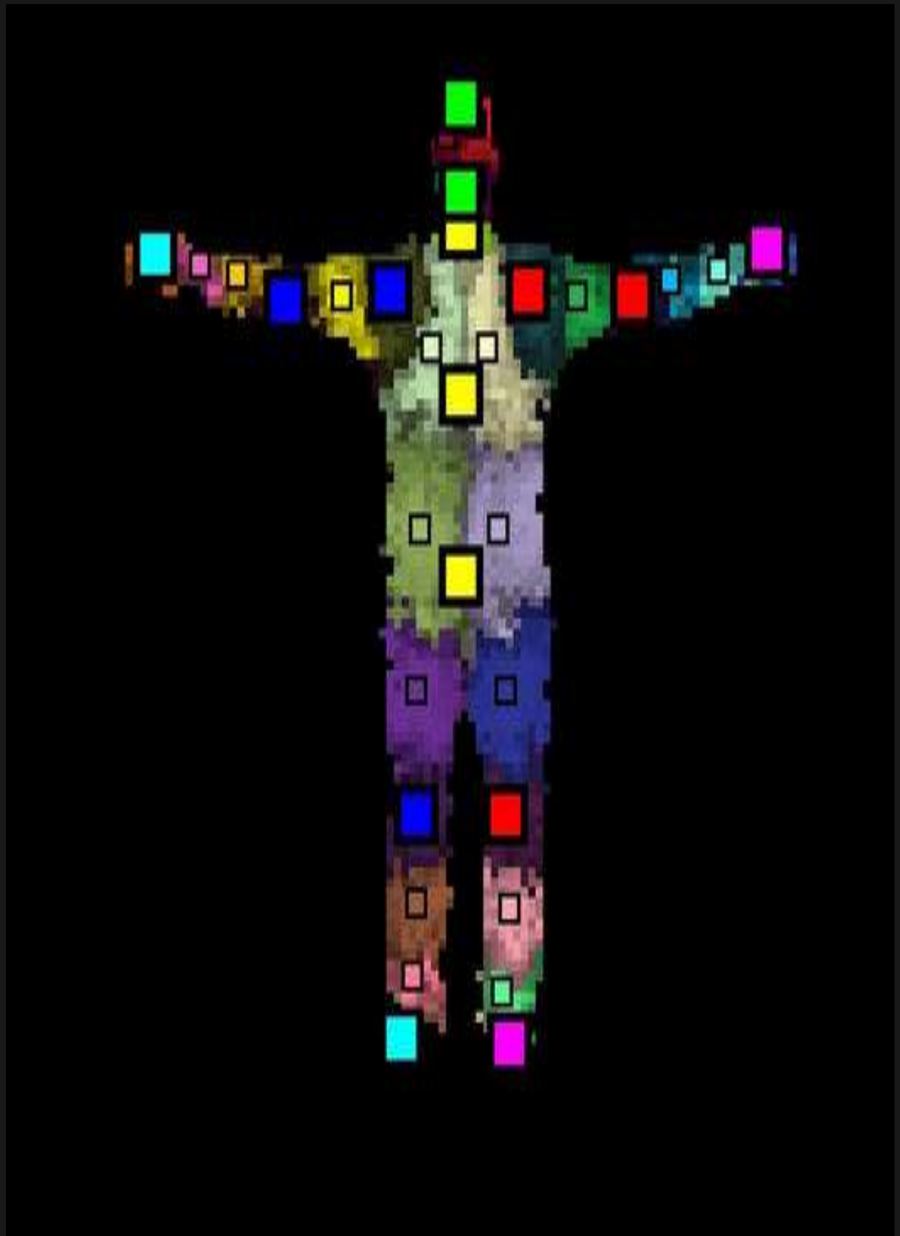
Buxton, W. (1990). A Three-State Model of Graphical Input. In D. Diaper et al. (Eds), *Human-Computer Interaction - INTERACT '90*. Amsterdam: Elsevier Science Publishers B.V. (North-Holland), 449-456.

Pinching Detection

Wilson, A. Robust Vision-Based Detection of Pinching for One and Two-Handed Input, *UIST 2006.*

Skeleton tracking (Kinect)

Skeletal Tracking - How it works



1. Classify each pixel's probability of being each of 32 body parts
2. Determine probabilistic cluster of body configurations consistent with those parts
3. Present the most probable pose to the user

Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore, [Real-time human pose recognition in parts from single depth images], in *Communications of the ACM (CACM)*, ACM, January 2013.

Hand Tracking

Leap Motion Controller:
2 IR cameras
3 IR LEDs

Alternate approach…

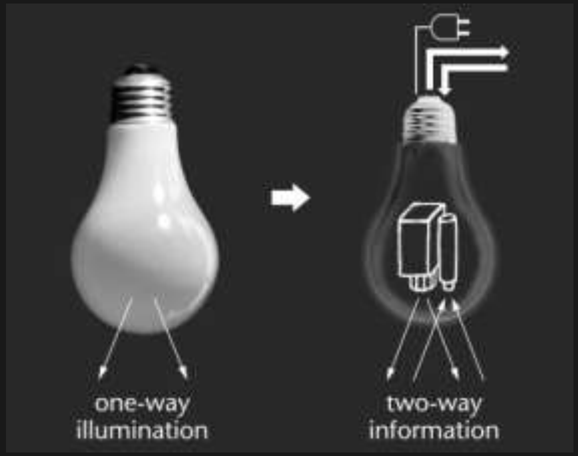
Use all the information from the camera.
Do not require a reduced representation.
Make it analog to real world.

Kruger, Videoplace '85

LightSpace

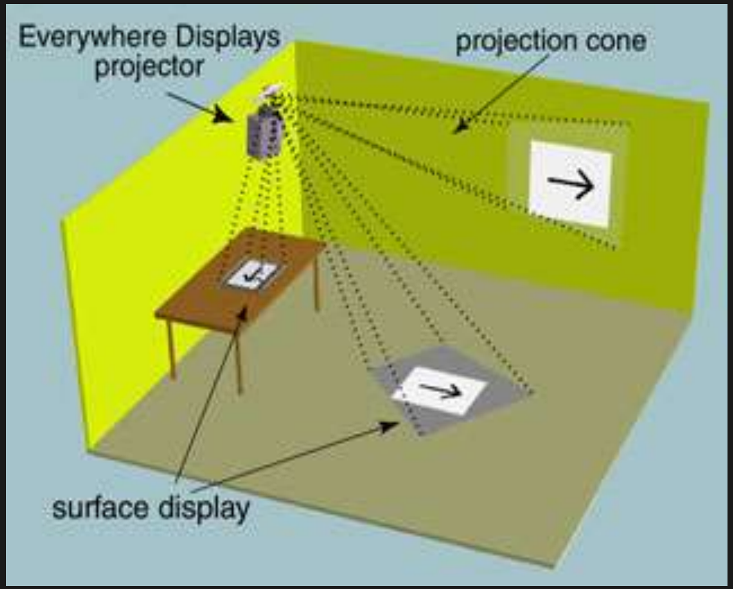
UIST 2010

Related



Underkoffler & Ishii, CHI '98


Raskar et al., SIGGRAPH '98



Pinhanez, UBICOMP '01

TOUCH ON EVERY SURFACE

Surface determined empirically

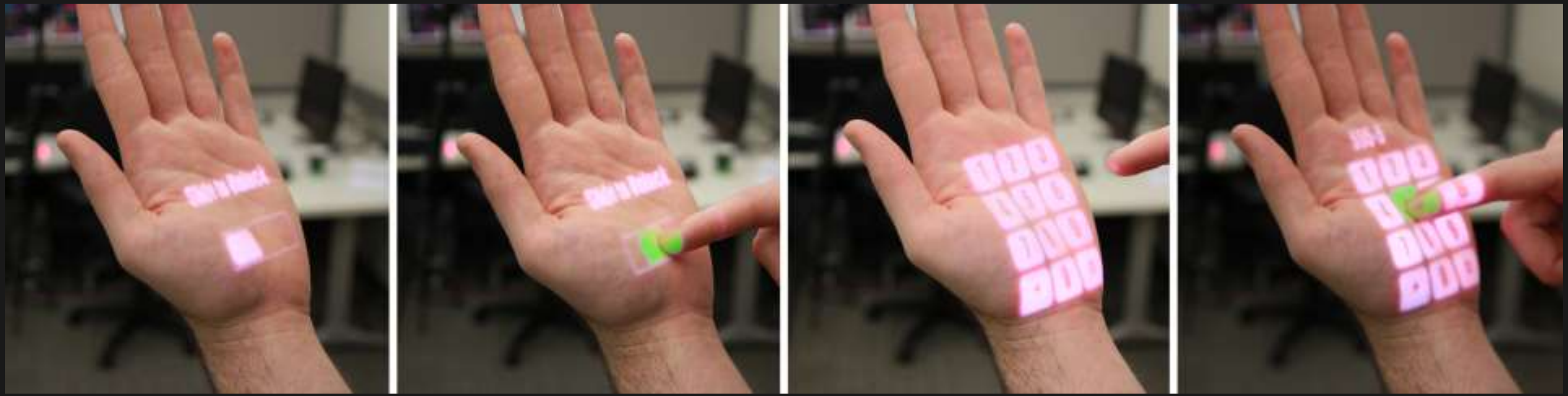
At 0.75m ~6mm

At 1.5m ~30mm

Wilson, ACM ITS 2010

But this works for static surfaces only!

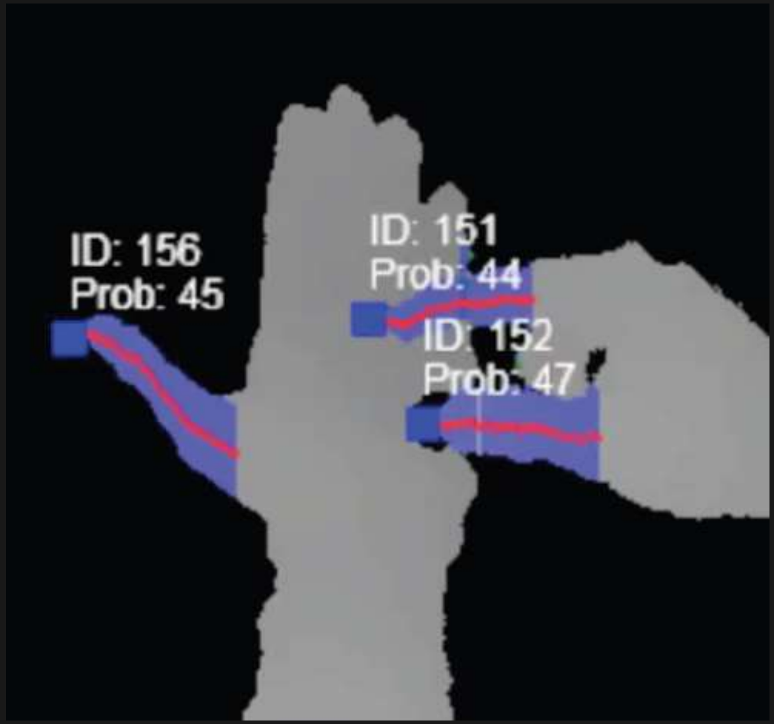
What about dynamic surfaces?

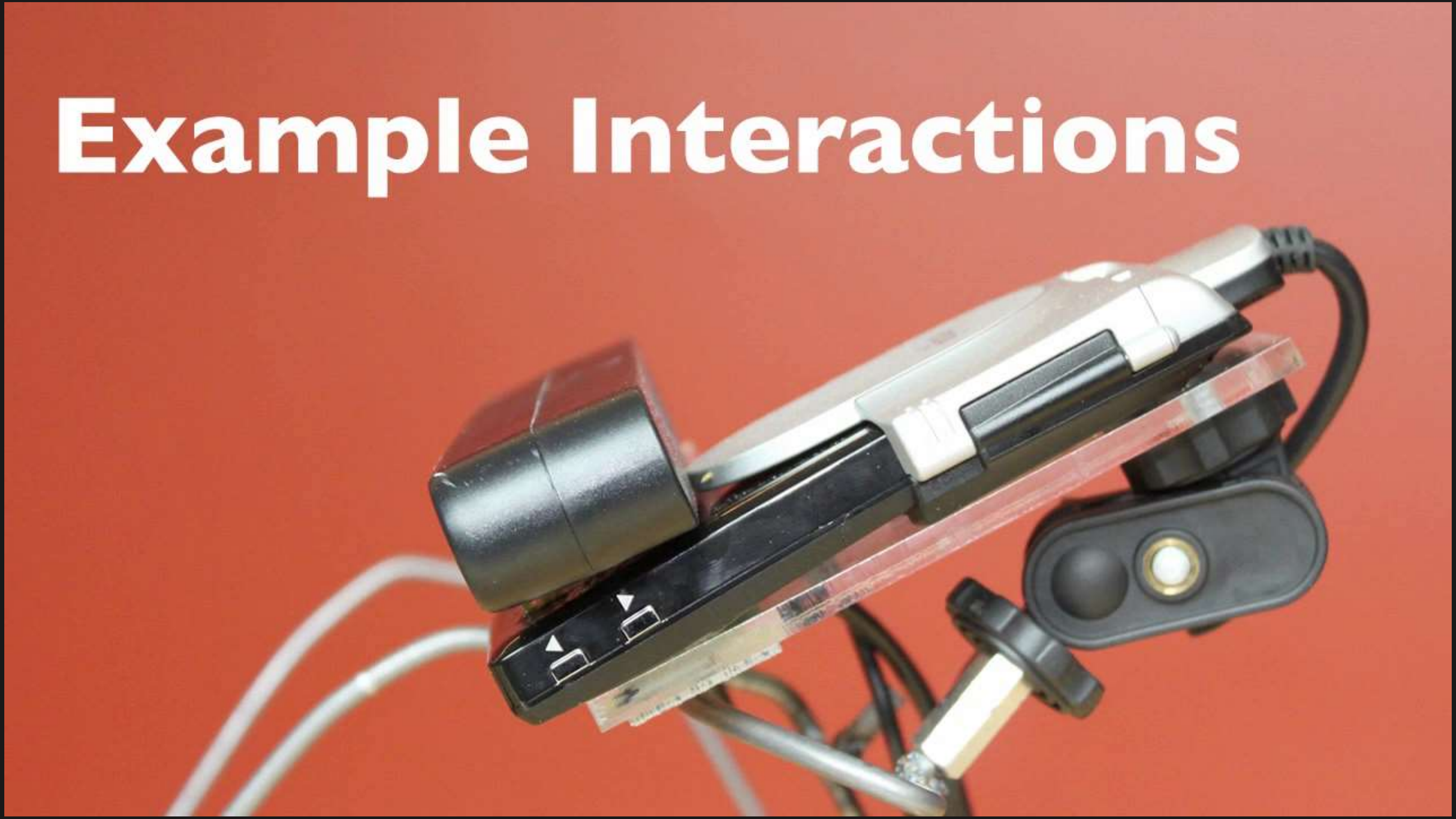


Harrison, Benko, and Wilson, ACM UIST 2011

Tracking high-level constructs (fingers)

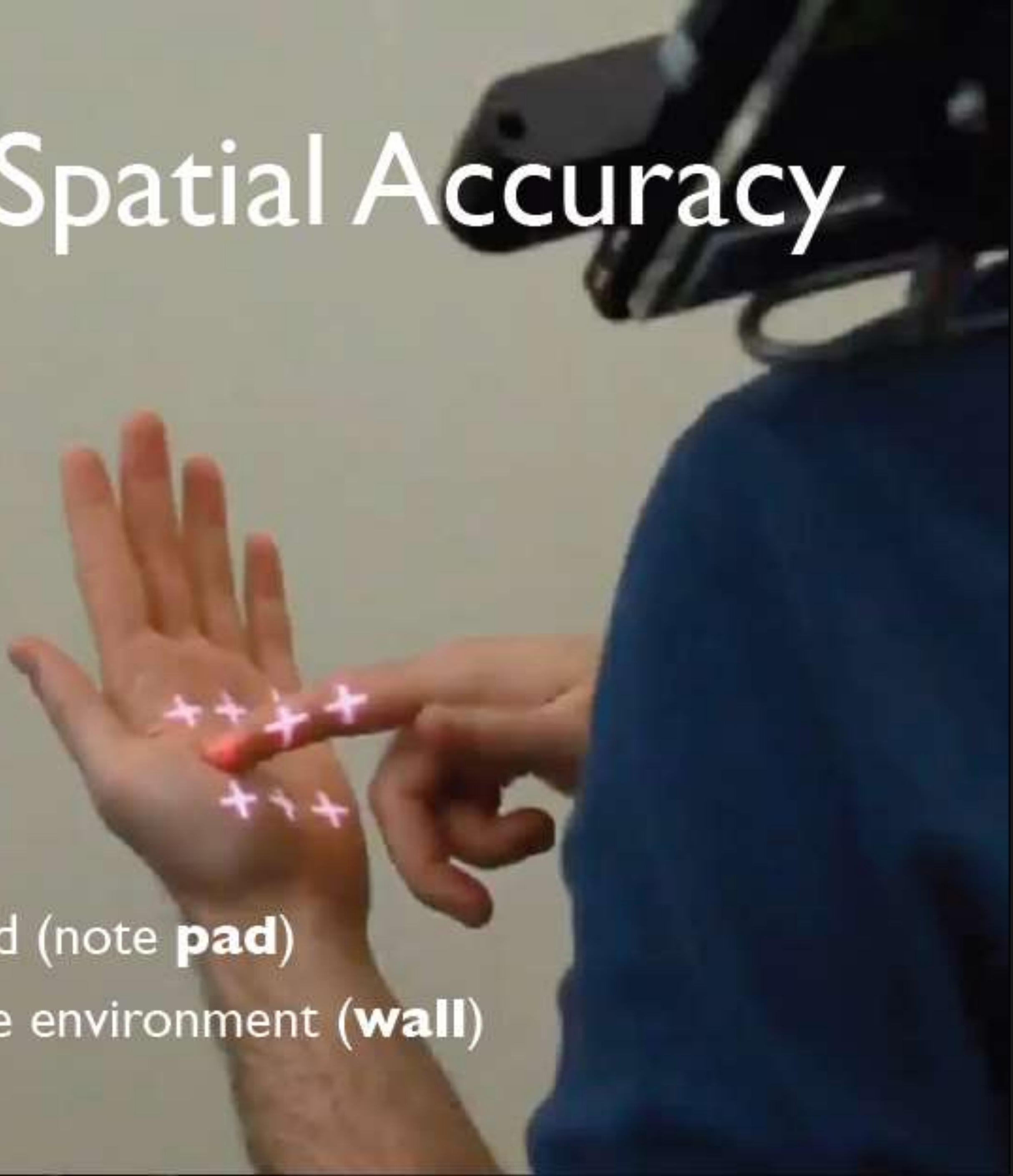
- Take only the ends of objects with physical extent (“fingertips”)
- Detect contact (“click”)
- Refinement of position while clicked (“drag”)





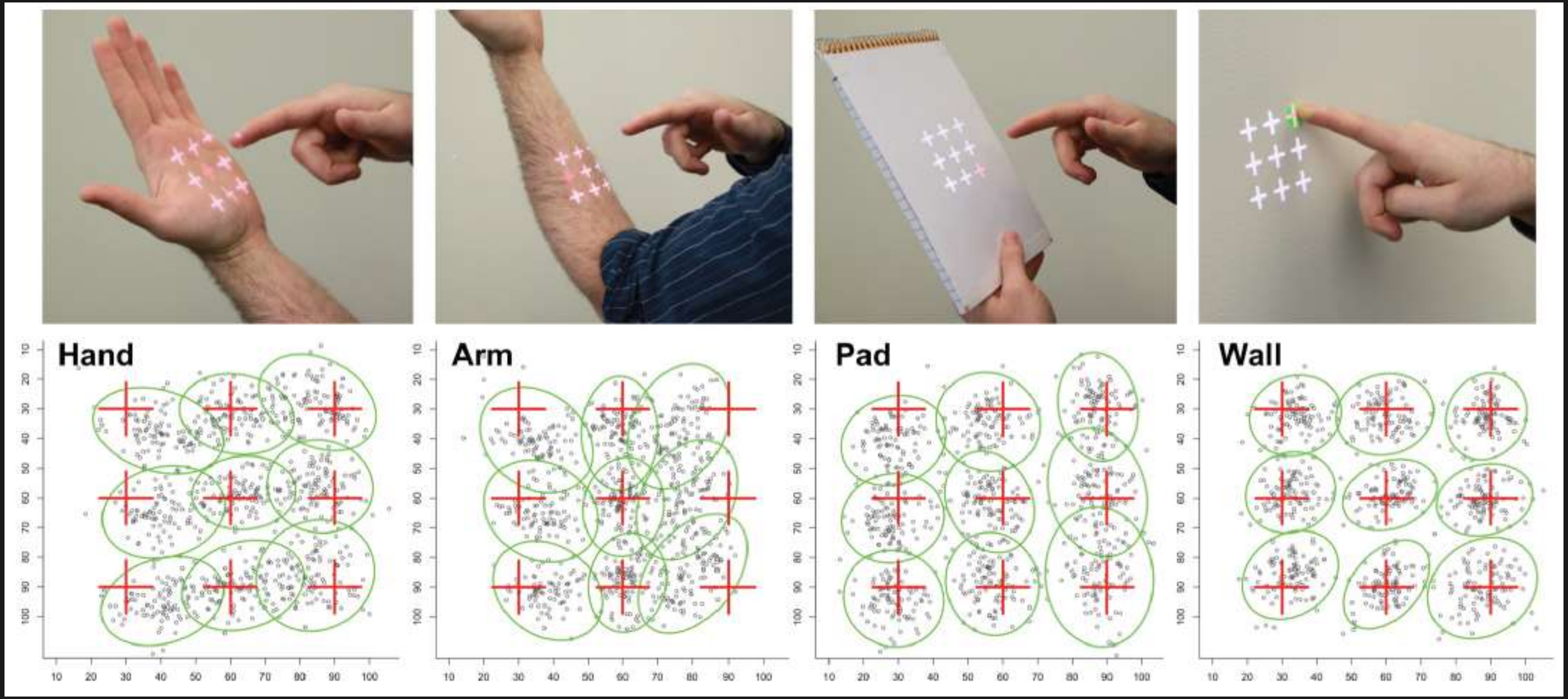
Harrison, Benko, and Wilson, ACM UIST 2011

“Click” Spatial Accuracy



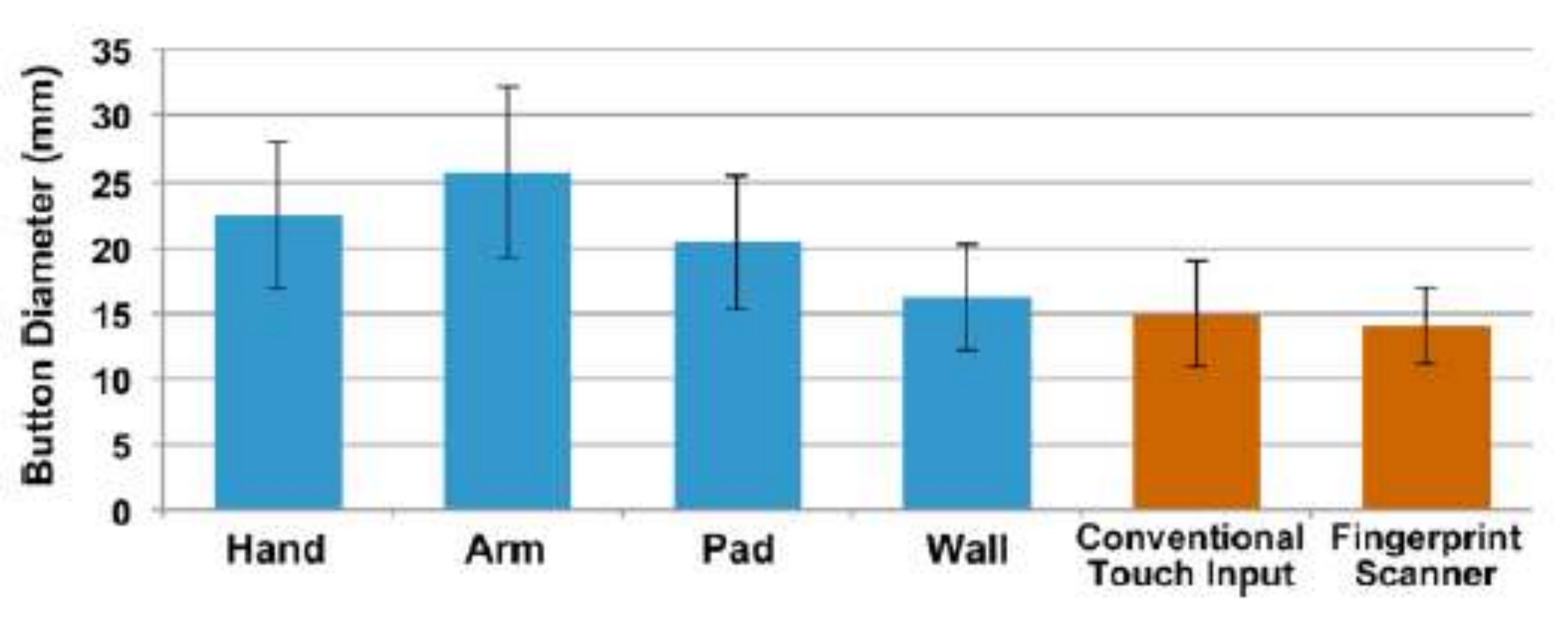
Four test surfaces:

- On body (**hand**)
- Object held in hand (note **pad**)
- Fixed surface in the environment (**wall**)
- Also added **arm**



6048 click trials

Click Spatial Accuracy



Holz and Baudisch - CHI '10

With 0.5s timeout rejection ~ 98.9% click accuracy

INTERACTING WITH 3D OBJECTS

Interactions in the Air

Hilliges, O., Izadi, S., Wilson, A., Hodges, S., Butz, A., and Garcia-Mendosa, A., UIST 2009

How to see a virtual 3D object in your hand?

How to manipulate it using the full dexterity of your hand?

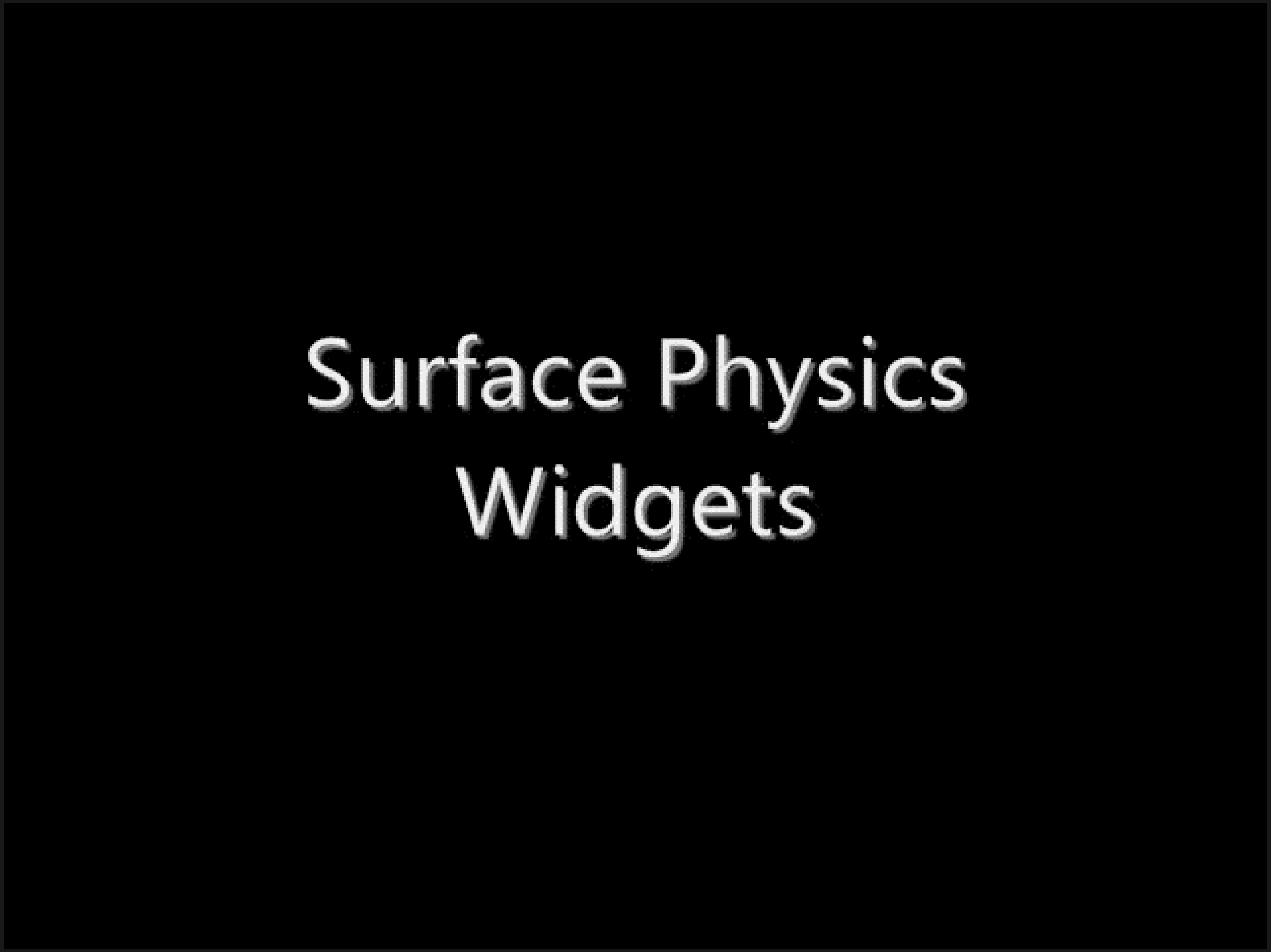
MirageTable

Benko, Jota, Wilson, CHI 2012

Discussion

- How do you evaluate the quality of a prototype system? (Does it even make sense to do so?)
- What are the limitations of using physics engine to drive interactions?

Surface Physics Widgets



Challenges with depth cameras

Hands are deformable

- Dynamic meshes are not supported

Depth cameras do not give you lateral forces

- Can’t place torque on an object

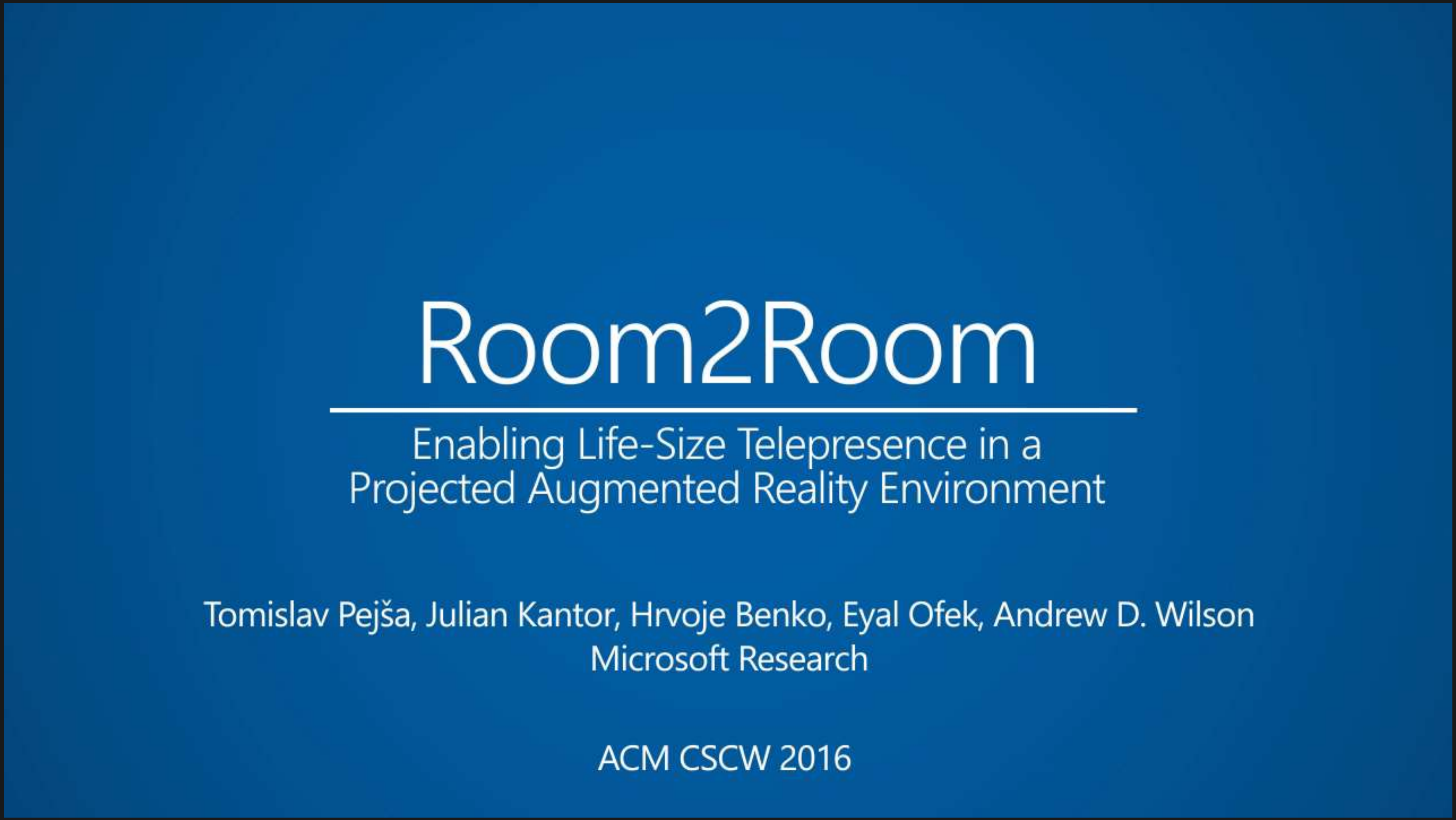
Lack of force feedback

- Grasping is tricky

Discussion

- How do you evaluate the quality of a prototype system? (Does it even make sense to do so?)
- What are the limitations of using physics engine to drive interactions?
- What the teleconference in the future look like? VR? AR?

Chasing Richer Telepresence



Pejsa, T., Kantor, J., Benko, H., Ofek, E., and Wilson, A. Room2Room CSCW 2016.

SENSOR FUSION: CAMERAS + OTHER SENSORS

Multimodal Sensor Fusion

Put that There (Bolt et al. 1980)

CrossMotion

CrossMotion

Fusing Device and Image Motion for User Identification, Tracking and Device Association

Andrew D. Wilson and Hrvoje Benko
Microsoft Research

ACM ICMI 2014

Discussion

- How “accurate” is CrossMotion?
- What are the limitations?
- What is the difference between tracking and identification?
- What makes this approach practical? Impractical?
- Can you think of other application areas particularly suited for CrossMotion?

MAKING SENSING SYSTEMS USABLE

Transition in How We Think About Interfaces

Interaction as Execution and Evaluation (Norman 1990)

Command line
GUI

Interaction as Conversation

Speech Interfaces
Ubiquitous Computing
Human Robot Interactions
Tangible Interfaces
Virtual/Augmented Reality

Conversations are not always specific

Potentially imprecise
Multimodal
Context is important
Not all options are visible/discoverable

Bellotti et al.: Five questions for designers of sensing systems

Address
Attention
Action
Alignment
Accident

Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. 2002. Making sense of sensing systems: five questions for designers and researchers. In Proceedings of CHI '02. http://doi.acm.org/10.1145/503376.503450

Bellotti et al: Five questions

Address:
- How do I address the system
- How do I address one (or more) of the many possible devices?
- How not to address the system?

GUI: keyboard, mouse, physical access

Sensing systems: Deal with signal vs. noise
- Magic keyword (“Alexa”, “Xbox”, “Siri”)
- Magic pose (Xbox: hand in front of the body)

Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. 2002. Making sense of sensing systems: five questions for designers and researchers. In Proceedings of CHI '02. http://doi.acm.org/10.1145/503376.503450

Bellotti et al: Five questions

Attention:

- How do I know the system is ready and attending to my actions?

GUI: graphical feedback, conventions (blinking cursor)

Sensing systems:

- What is appropriate feedback for mid-air interaction?

Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. 2002. Making sense of sensing systems: five questions for designers and researchers. In Proceedings of CHI '02. http://doi.acm.org/10.1145/503376.503450

Bellotti et al: Five questions

Action:

- How do I effect a meaningful action?
- How to control its extent?
- How to specify a target or targets for my action?

GUI: click to select, click+drag to multiple select, etc.

Sensing Systems:

- Selection mechanism?

Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. 2002. Making sense of sensing systems: five questions for designers and researchers. In Proceedings of CHI '02. http://doi.acm.org/10.1145/503376.503450

Bellotti et al: Five questions

Alignment:

- How do I know the system is doing (has done) the right thing?

GUI: feedback conventions, progress bars, drag+drop interactions, etc.

Sensing Systems: How to make system state perceivable, or query-able?

Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. 2002. Making sense of sensing systems: five questions for designers and researchers. In Proceedings of CHI '02. http://doi.acm.org/10.1145/503376.503450

Bellotti et al: Five questions

Accident:

- How do I avoid mistakes?

GUI: Direct manipulation, Undo, Delete, Preview actions.

Sensing Systems:
Unintended actions? How to undo? How to cancel action in progress?

Victoria Bellotti, Maribeth Back, W. Keith Edwards, Rebecca E. Grinter, Austin Henderson, and Cristina Lopes. 2002. Making sense of sensing systems: five questions for designers and researchers. In Proceedings of CHI '02. http://doi.acm.org/10.1145/503376.503450

SUMMARY

Camera sensing…

… enables high-bandwidth interactions!

… enables rich virtual/augmented reality!

… enables rapid interactive device prototyping!

… is easily combined with other sensors!

Hrvoje Benko
benko@microsoft.com
http://research.microsoft.com/~benko