Rich Augmented Reality Interactions

(without goggles, gloves or 3D trackers)

Hrvoje Benko

Researcher, MSR Redmond

Microsoft Research

Microsoft Surface

Microsoft Touch Mouse

THE DISCONNECT BETWEEN REAL AND DIGITAL WORLDS

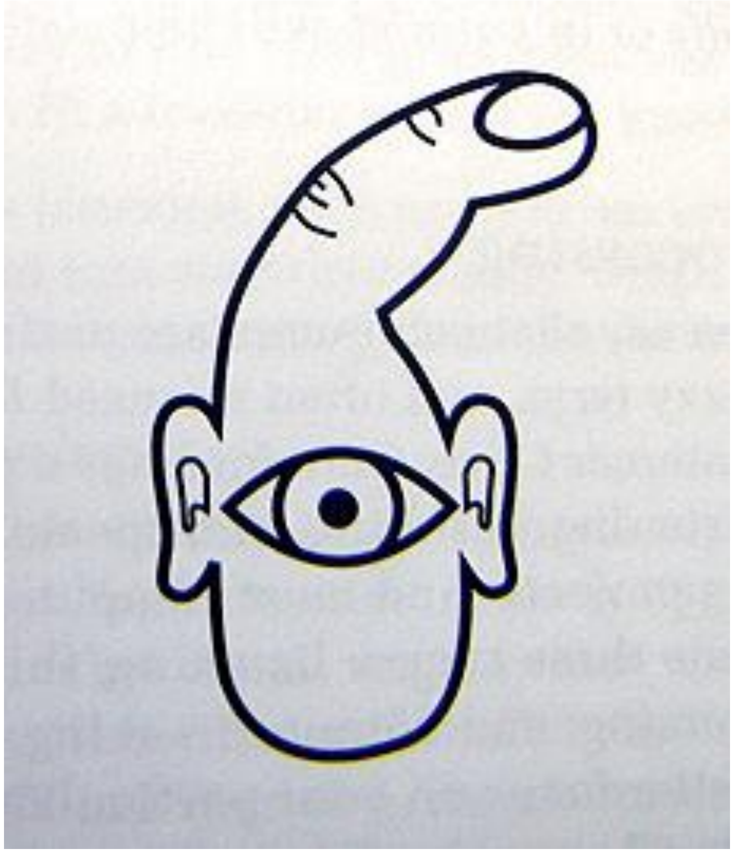
Visually rich

Call of Duty Modern Warfare 2

Visually rich

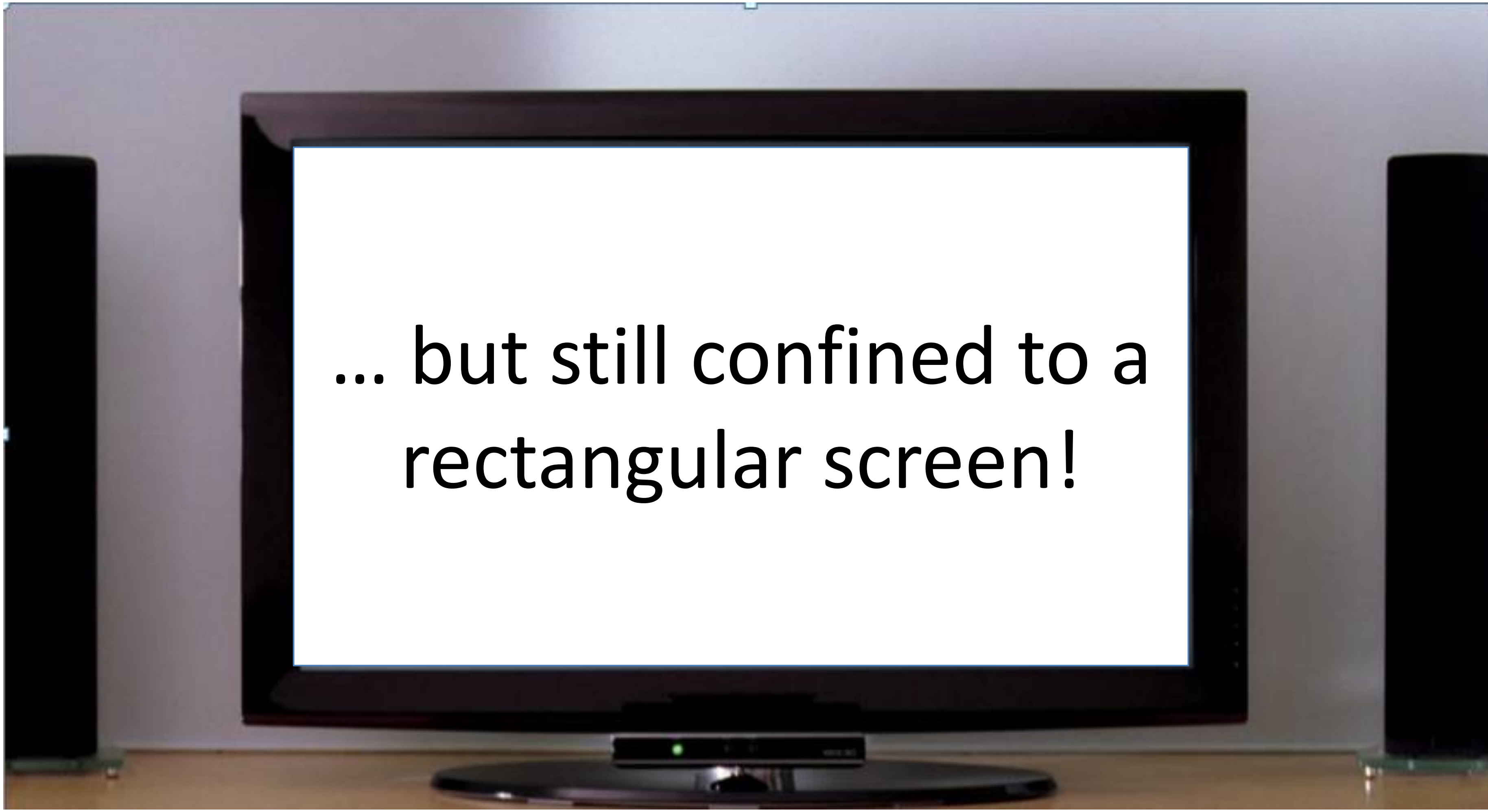
Interactively poor

How the computer sees us!



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

Kinect


… but still confined to a
rectangular screen!

REAL
WORLD

screen

DIGITAL
WORLD

An opportunity…

Depth camera

Projector

… to enable rich interactive experiences anywhere,

… to bridge the gap between “real” and “virtual” worlds.

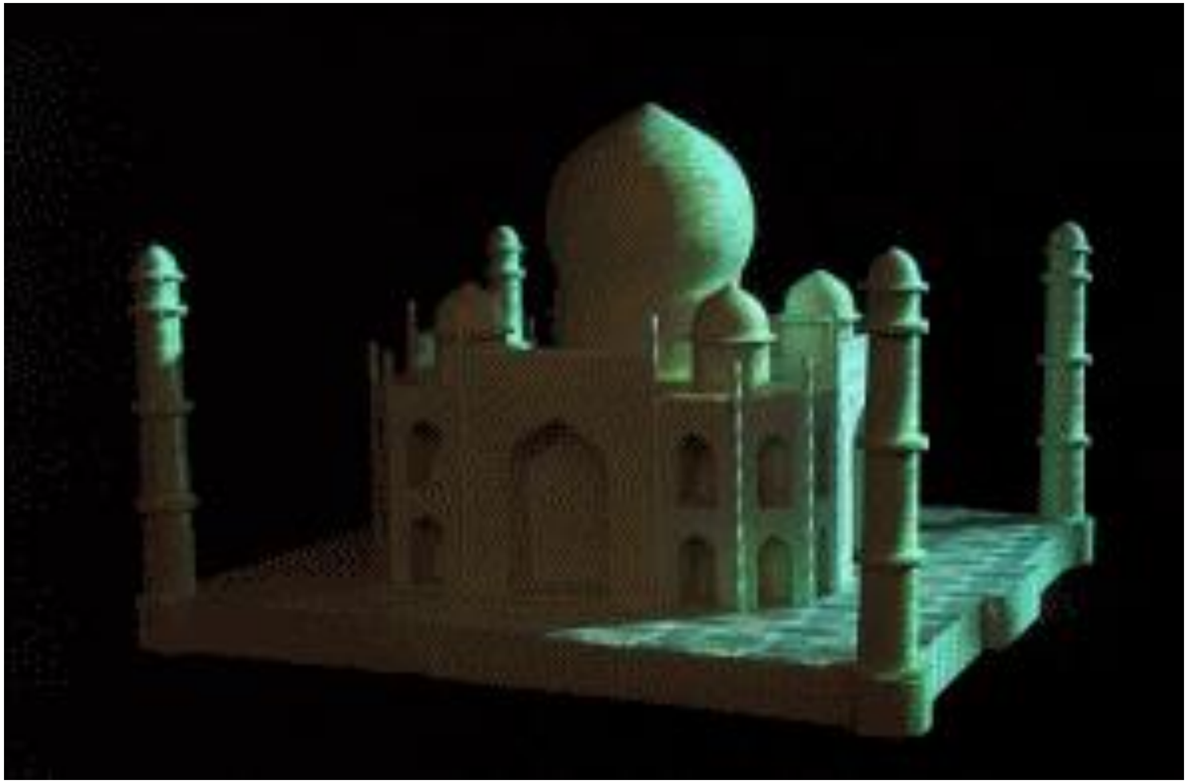
Augmented reality

Spatial

Deviceless

High-fidelity

Spatial Augmented Reality



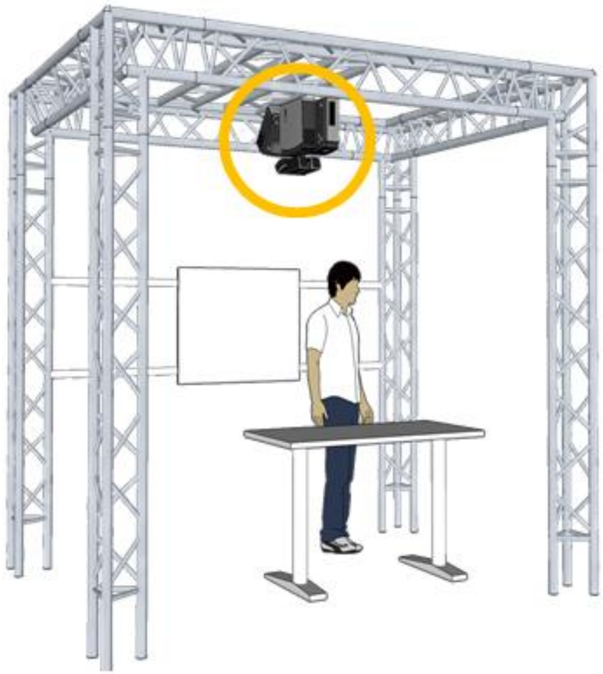
Raskar et al., EuroGraphics '01

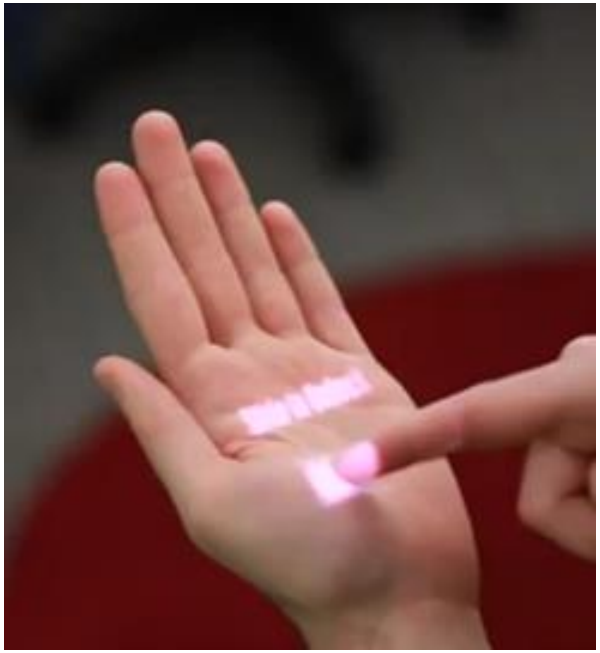
Microsoft Research

Interact with digital and real world

A. Wilson, ACM ITS 2007.

Depth camera + projector





INTERACTIVITY EVERYWHERE

TOUCH ON EVERY SURFACE

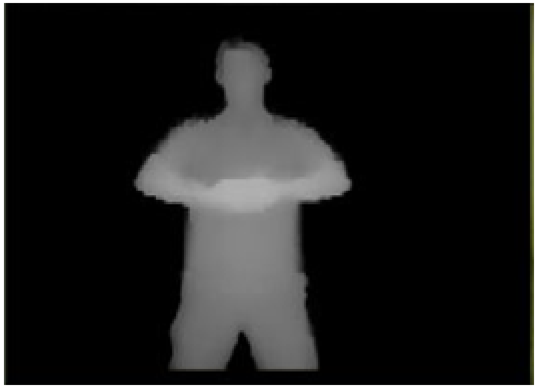
INTERACTING WITH 3D OBJECTS

Depth sensing cameras

Color + depth per pixel: RGBZ

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

- Stereo
- Time of flight
- Structured light



ENABLING INTERACTIVITY EVERYWHERE

Microsoft Research

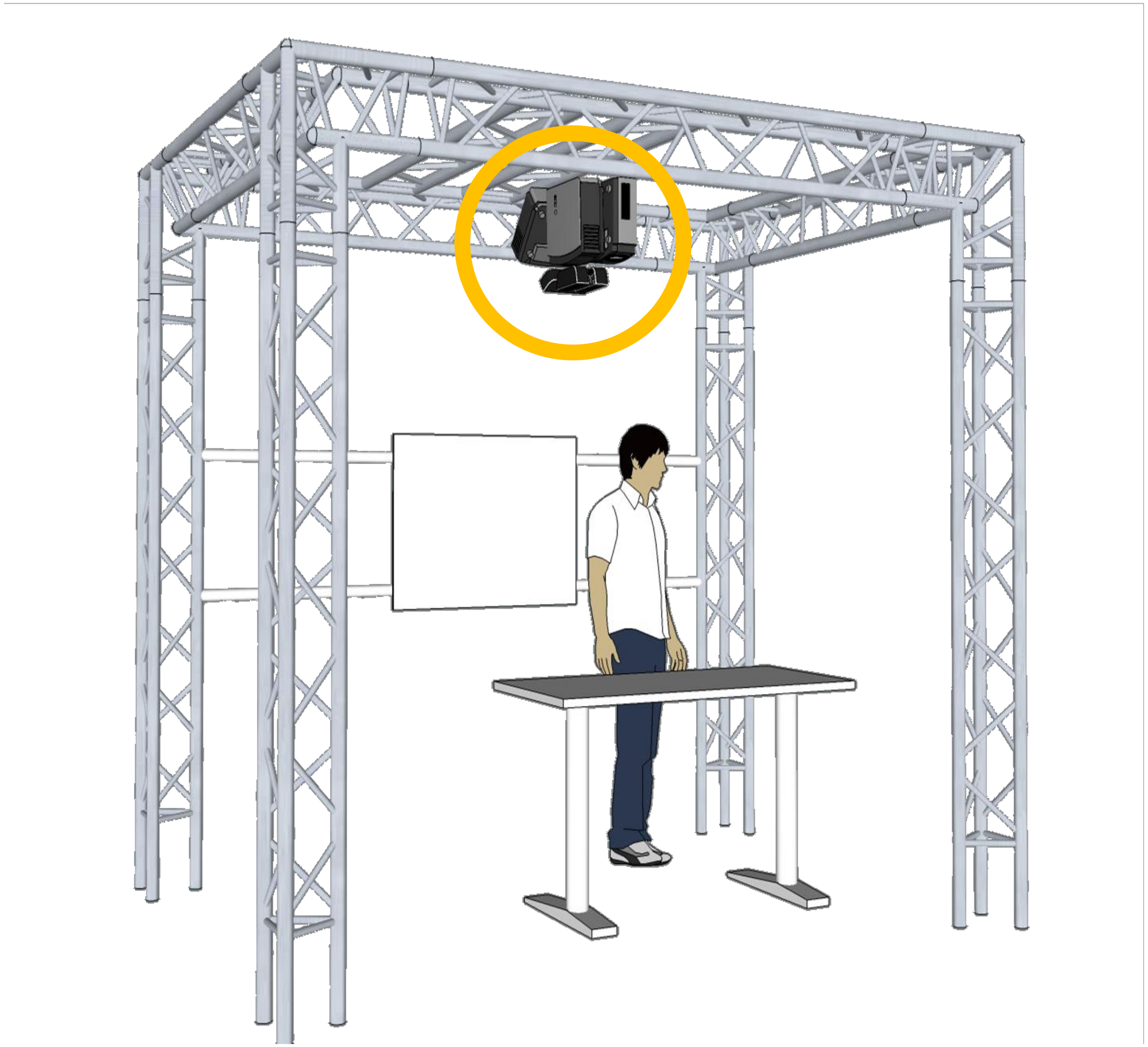
What if you could use any available surface (including your body) as an *interactive* surface?

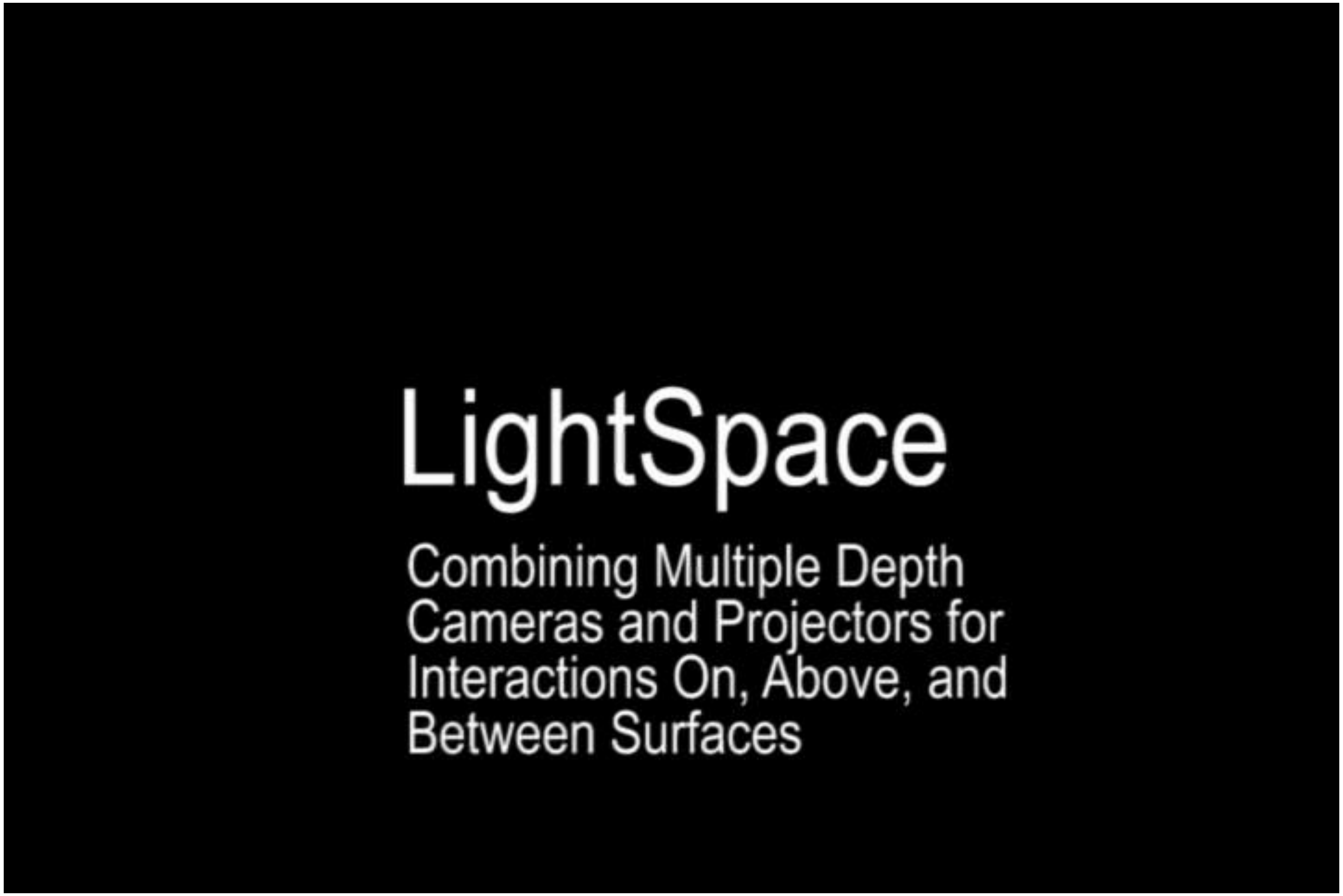
Rather than *reach* for a device…

simply *touch* where you want to see information…

and *interact* with it.

LightSpace

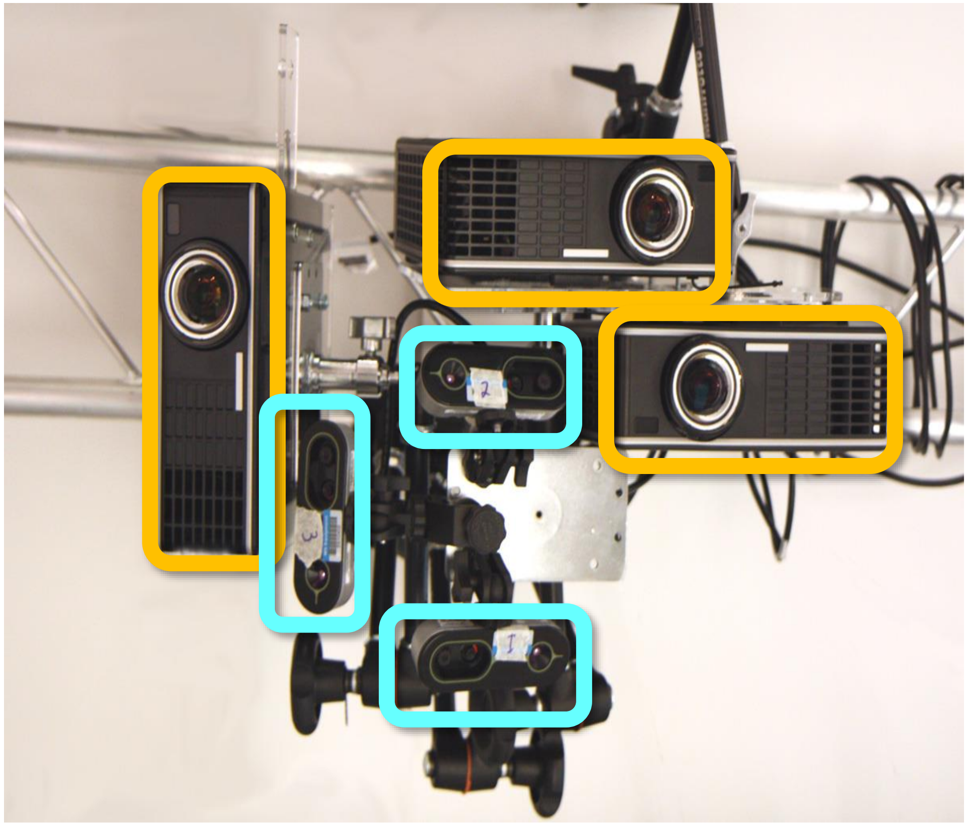




Wilson & Benko, ACM UIST, 2010

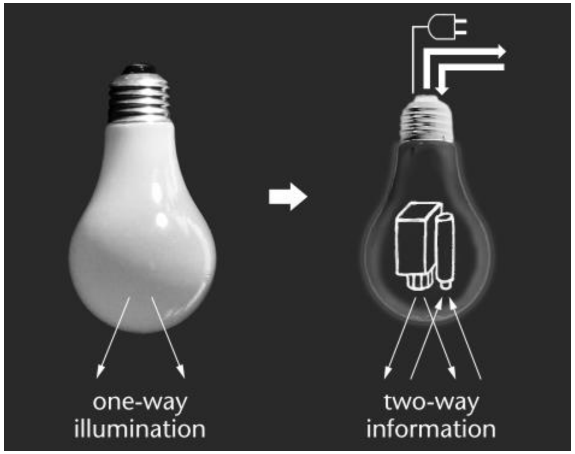
LightSpace Implementation

Projectors



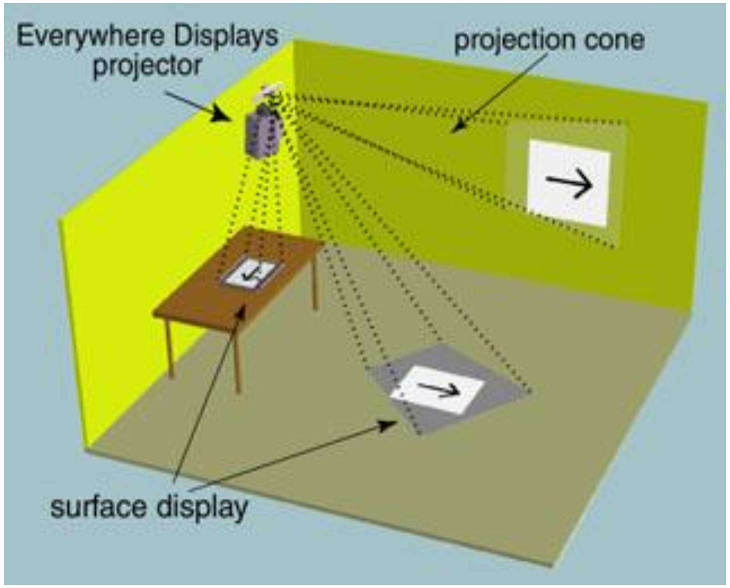
PrimeSense
Depth
Cameras

Related



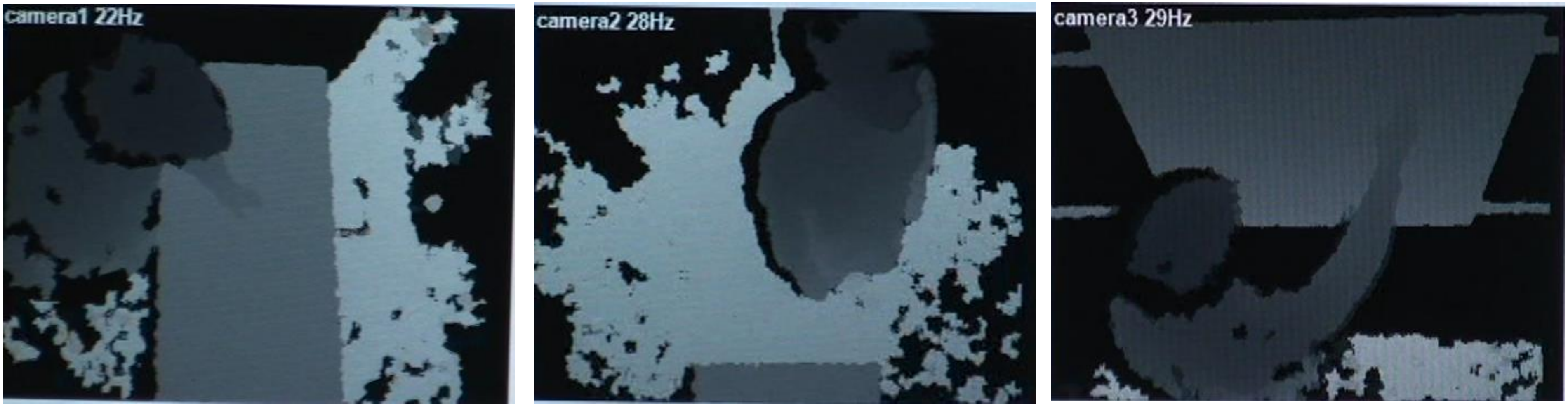
Underkoffler & Ishii, CHI ’98

Raskar et al., SIGGRAPH ‘98



Pinhanez, UBICOMP ‘01

PrimeSense depth cameras



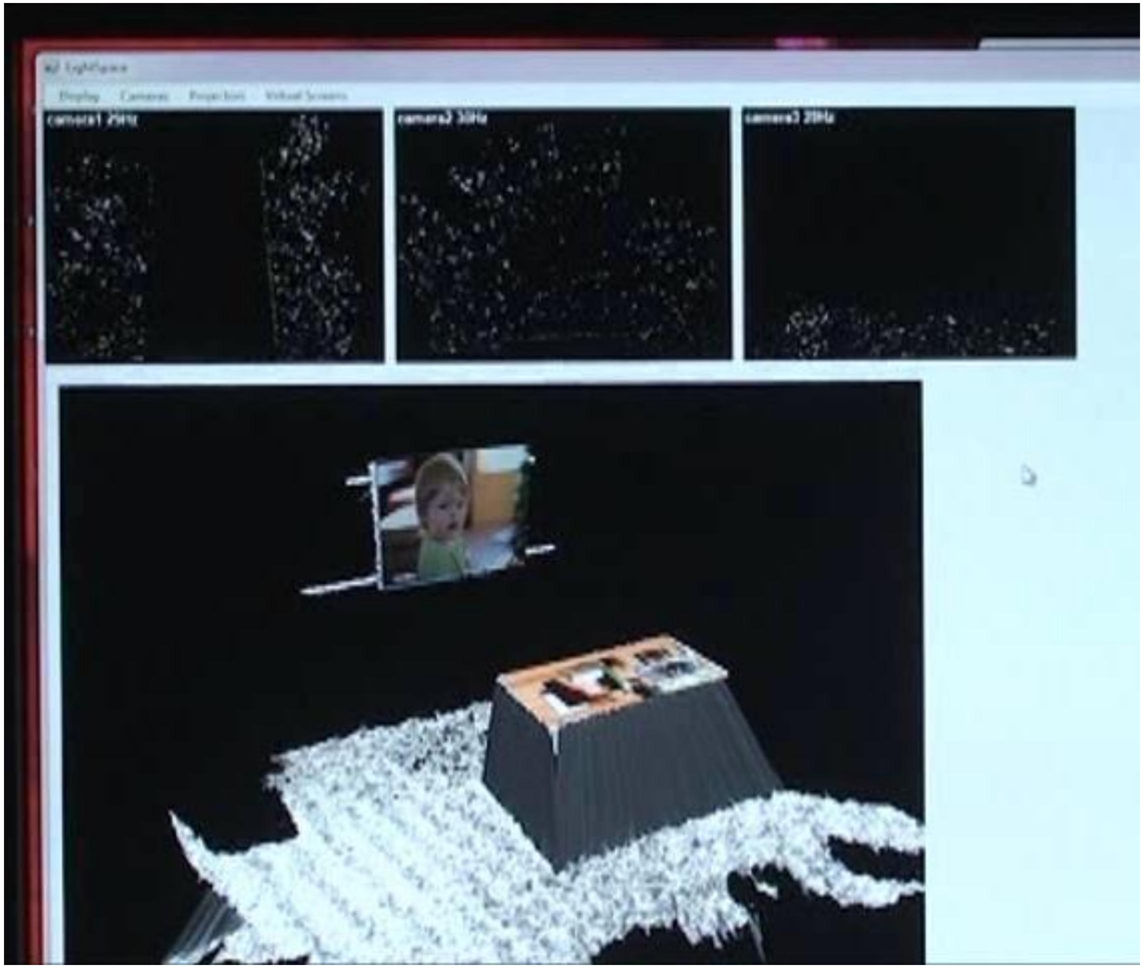
320x240 @ 30Hz

Depth from projected structured light

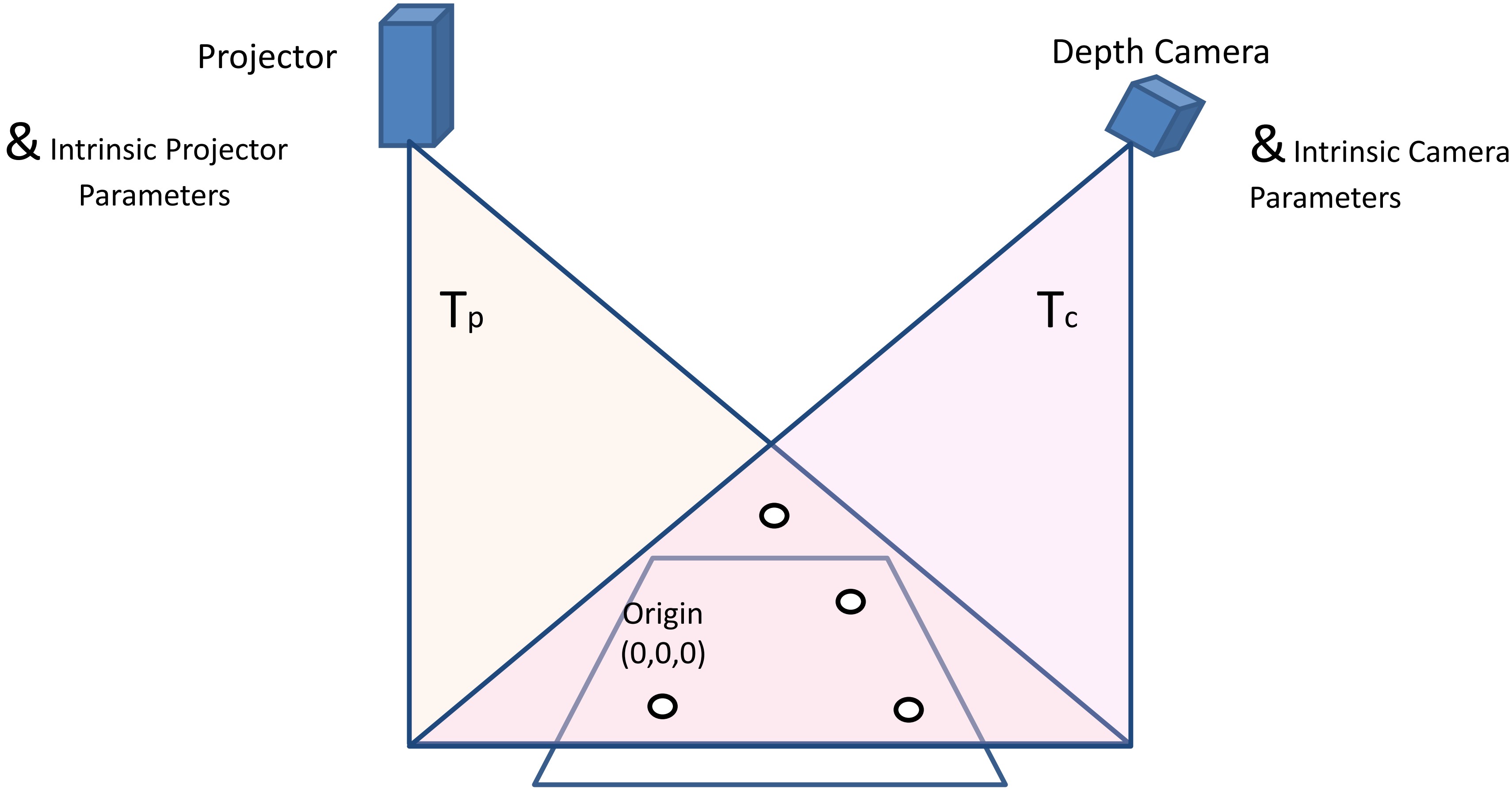
Small overlapping areas

Extended space coverage

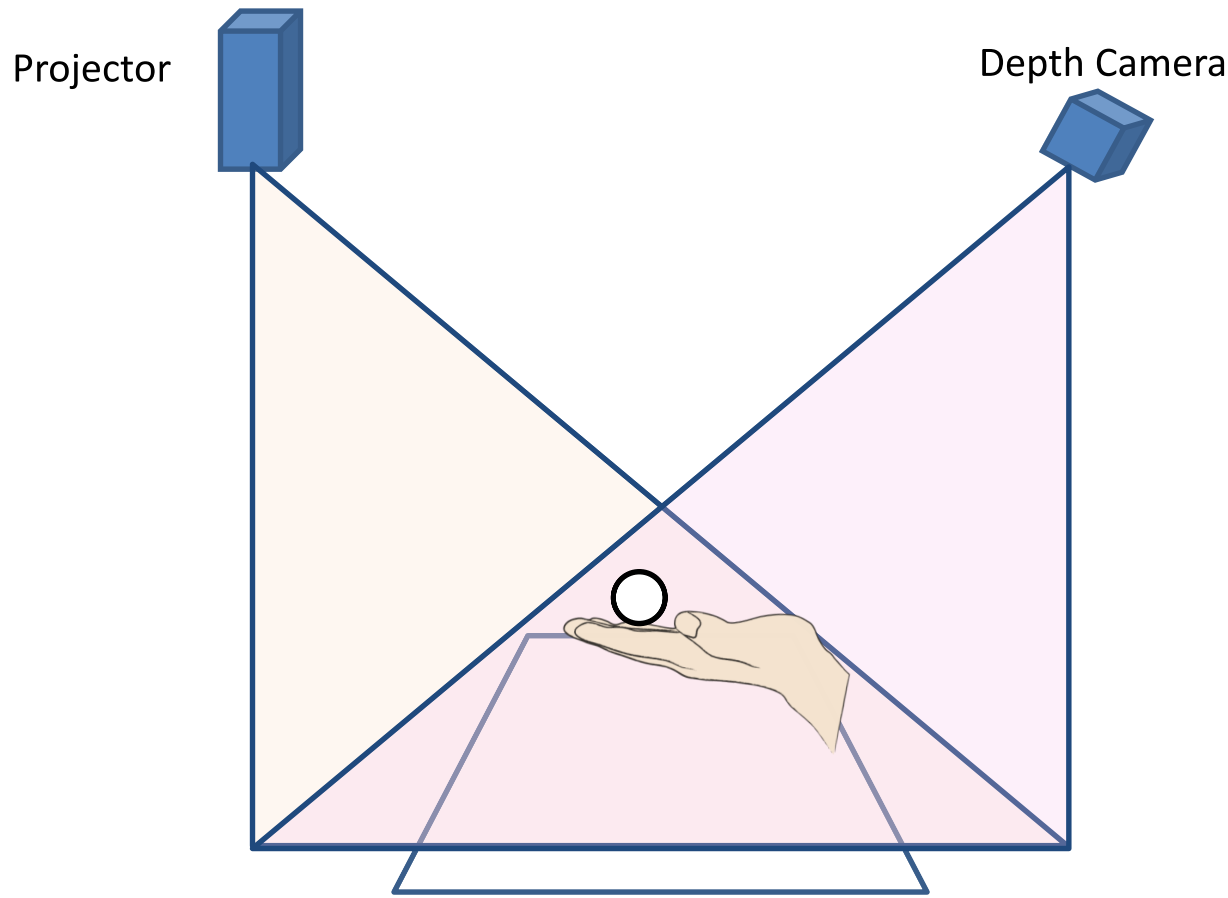
Unified 3D Space



Camera & projector calibration



Camera & projector calibration



LightSpace authoring

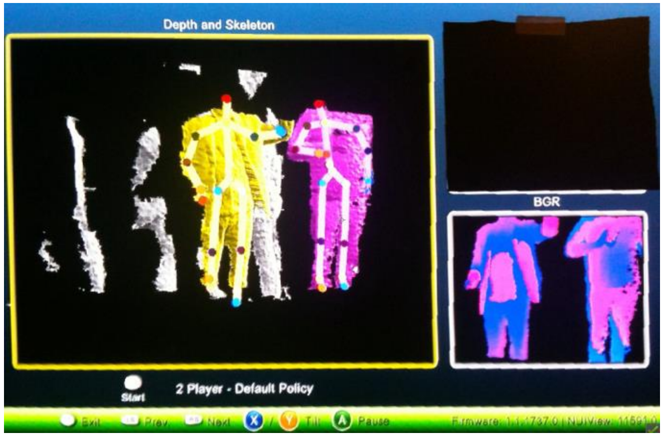
All in real world coordinates.

Irrespective of “which” depth camera.

Irrespective of “which” projector.

LIGHTSPACE INTERACTIONS

Skeleton tracking (Kinect)



Our approach

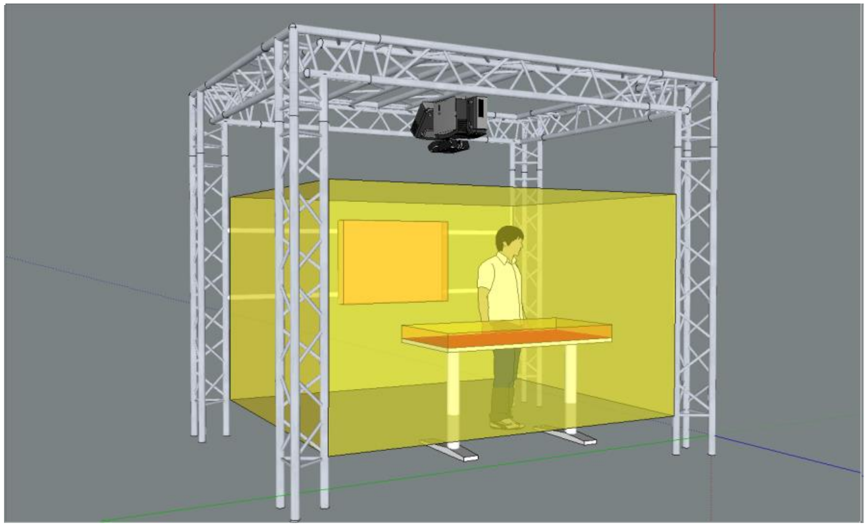
Use the full 3D mesh.

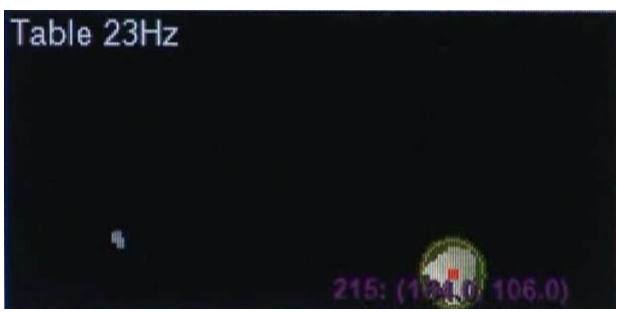
Preserve the analog feel through physics-like behaviors.

Reduce the 3D reasoning to 2D projections.

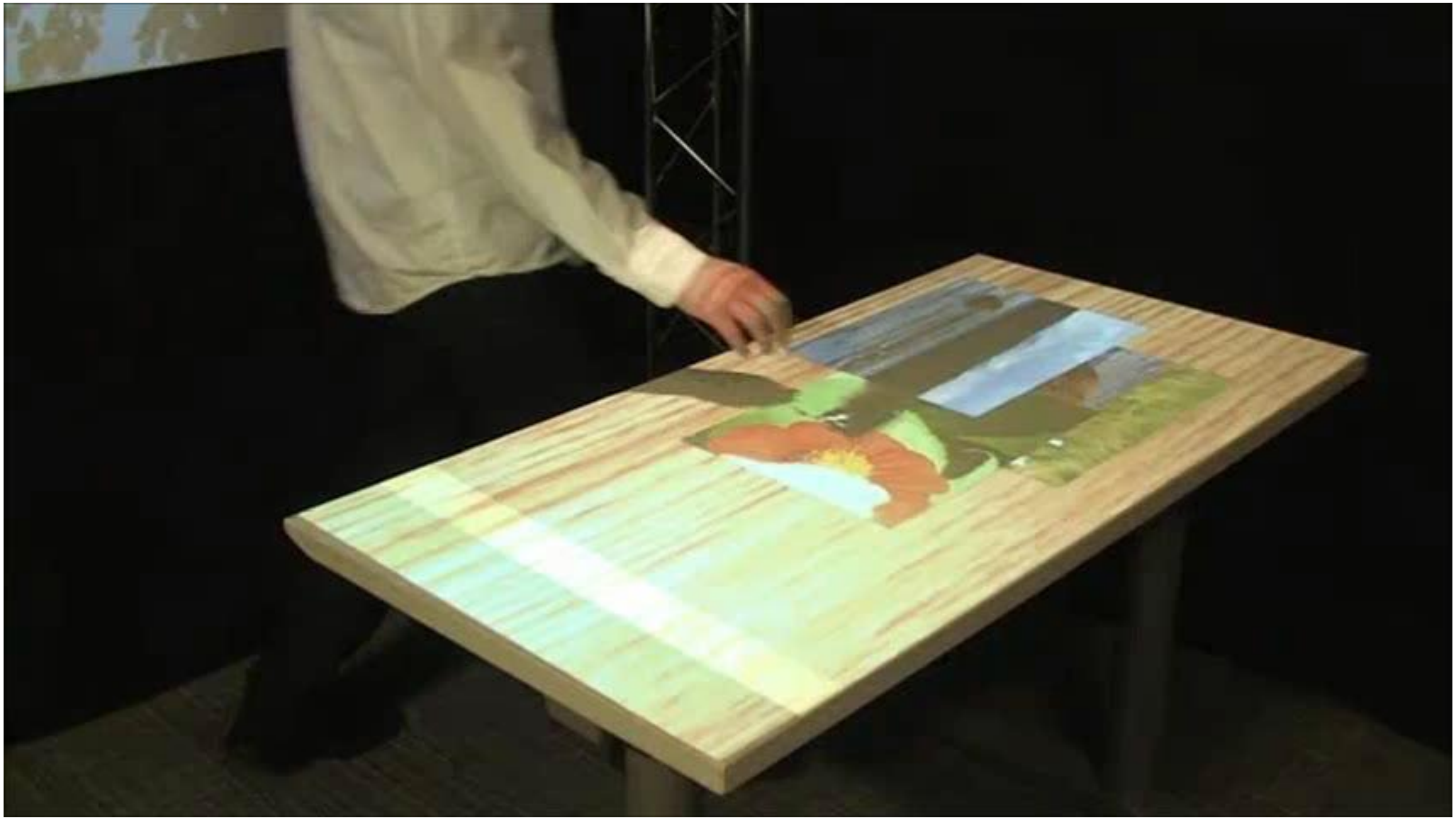
Pseudo-physics behavior

Virtual depth cameras





Simulating virtual surfaces



Through-body connections

Microsoft Research

Physical connectivity

What is missing?

LightSpace

- “Touches” are hand blobs
- All objects are 2D
- Very coarse manipulations

Ideally

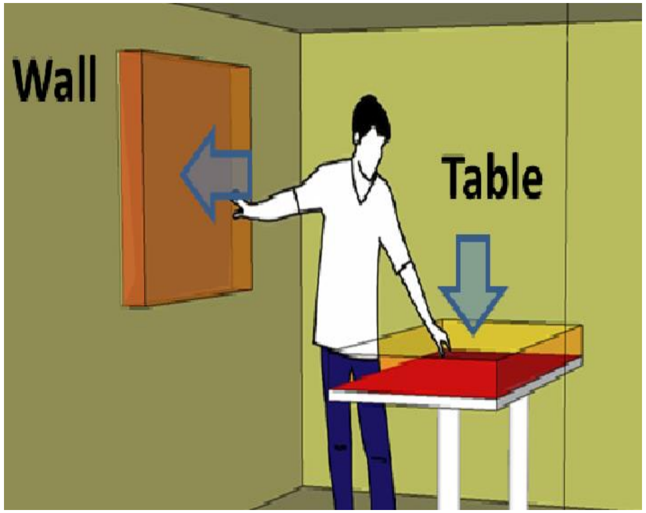
- Multi-touch
- 3D virtual objects
- Full hand manipulations

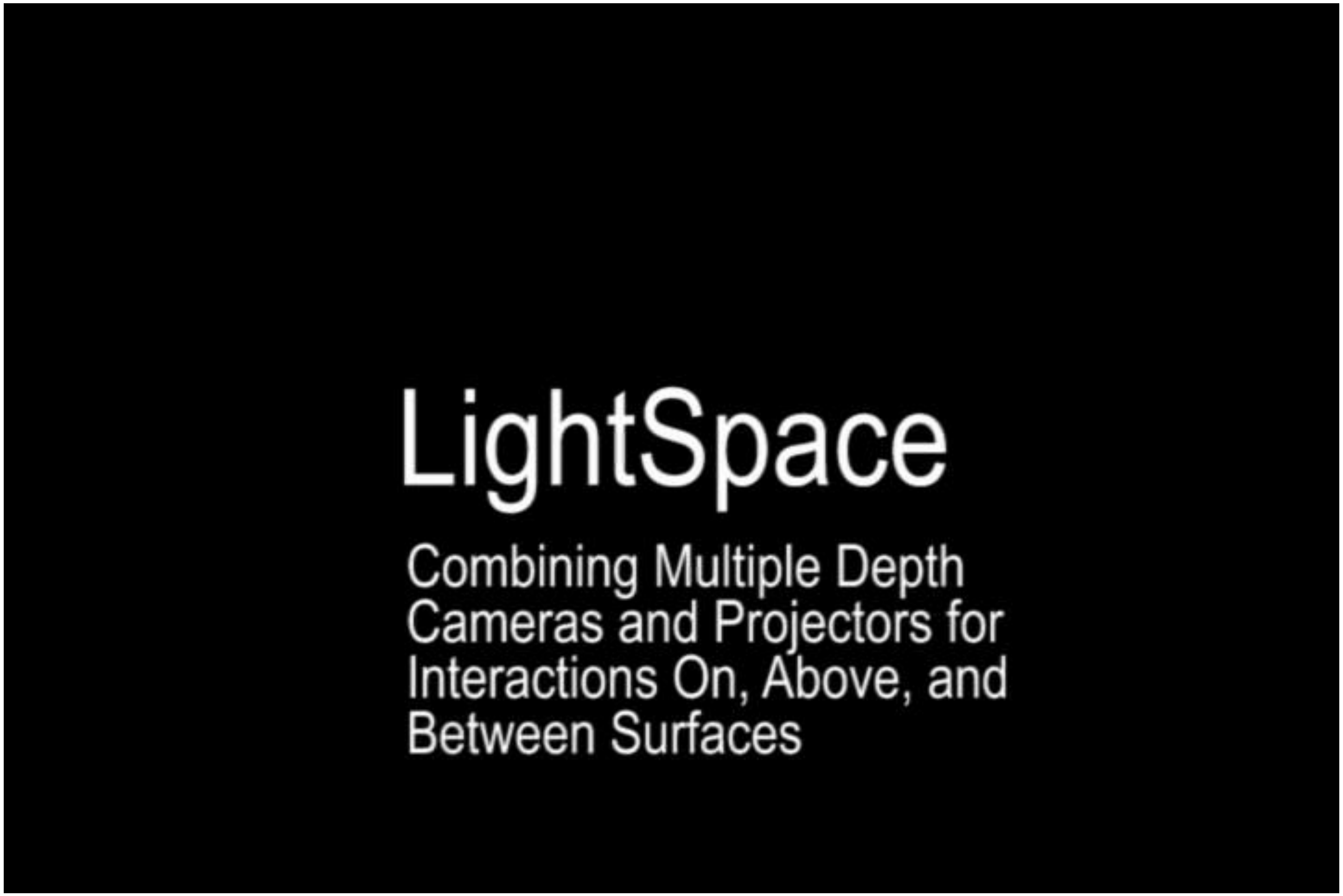
TOUCH ON EVERY SURFACE

How to get the surface?

Analytically

- Problems:
  - Slight variation in surface flatness
  - Slight uncorrected lens distortion effect in depth image
  - Noise in depth image





Wilson & Benko, ACM UIST, 2010

How to get the surface?

Empirically

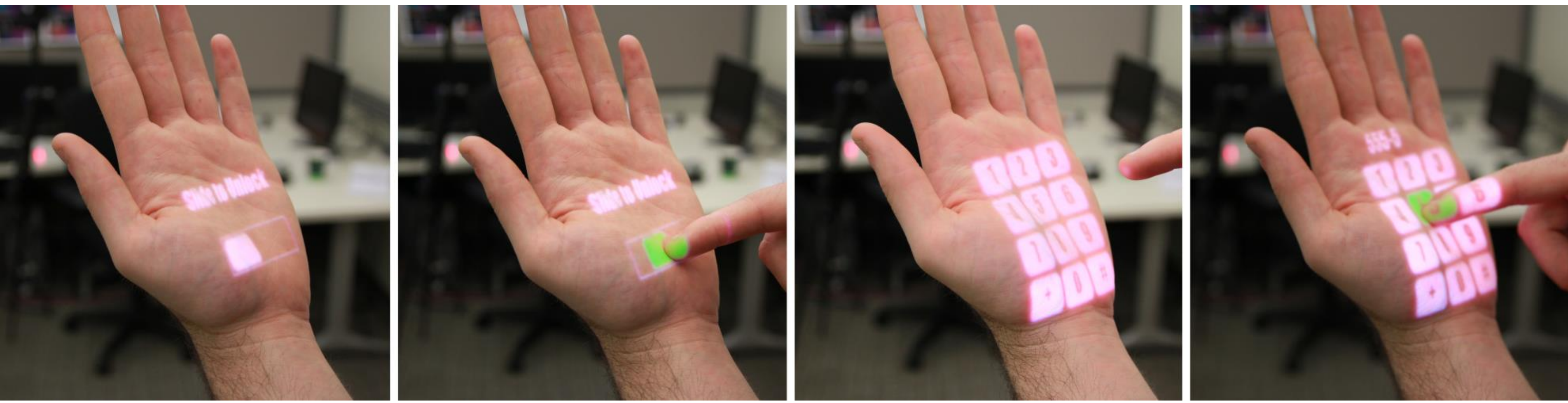
- Take per-pixel statistics of the empty surface
  - Can accommodate different kinds of noise
  - Can model non-flat surfaces
- Observations:
  - Noise is not normal, nor the same at every pixel location
  - Depth resolution drops with distance

Surface determined empirically

Wilson, ACM ITS 2010

But this works for
static surfaces only!

What about dynamic surfaces?



~~How to get the surface?~~

~~What is a surface?~~

Can we track the finger?

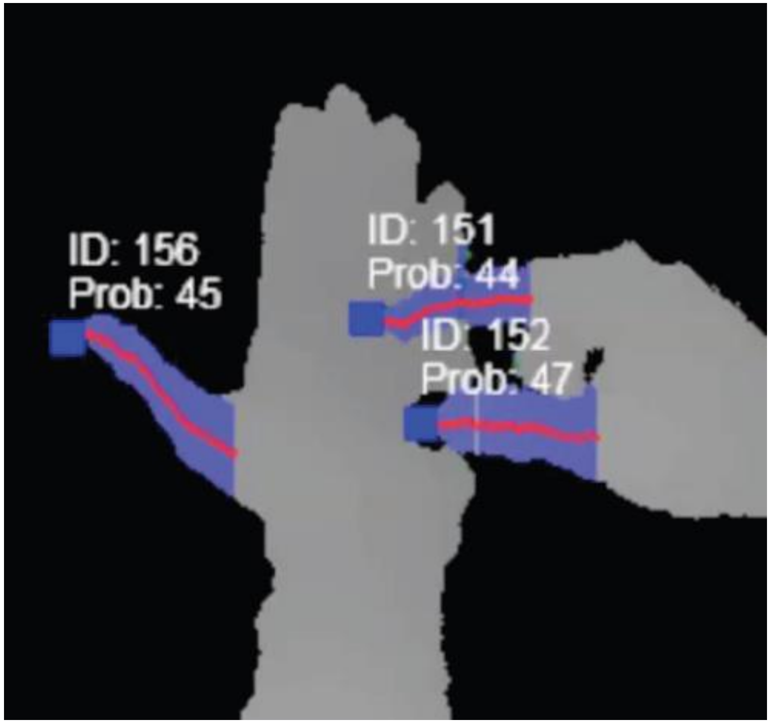
- Hard in general
- Simple from a body-centric perspective

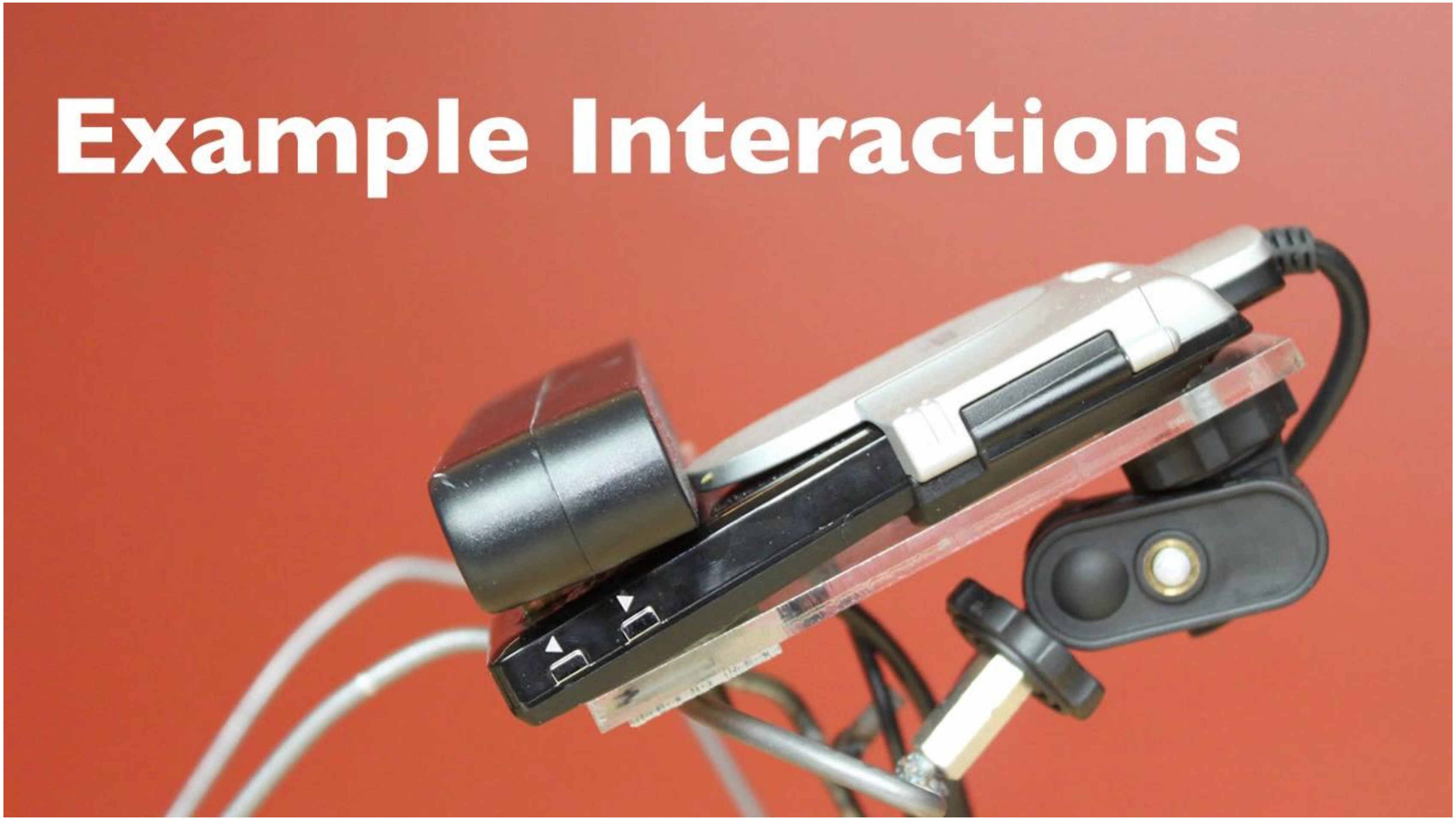
Microsoft
Research

Harrison, Benko, and Wilson, ACM UIST 2011

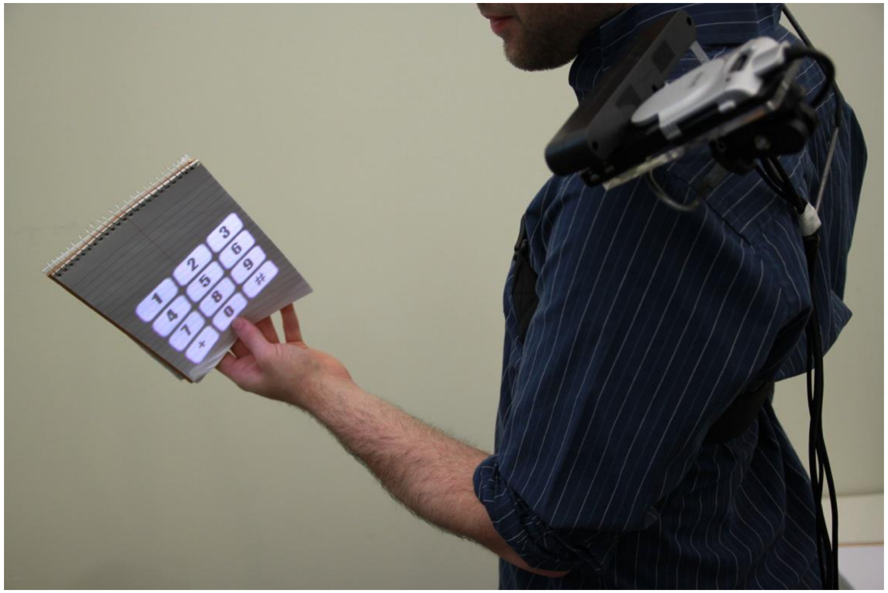
Tracking high-level constructs (fingers)

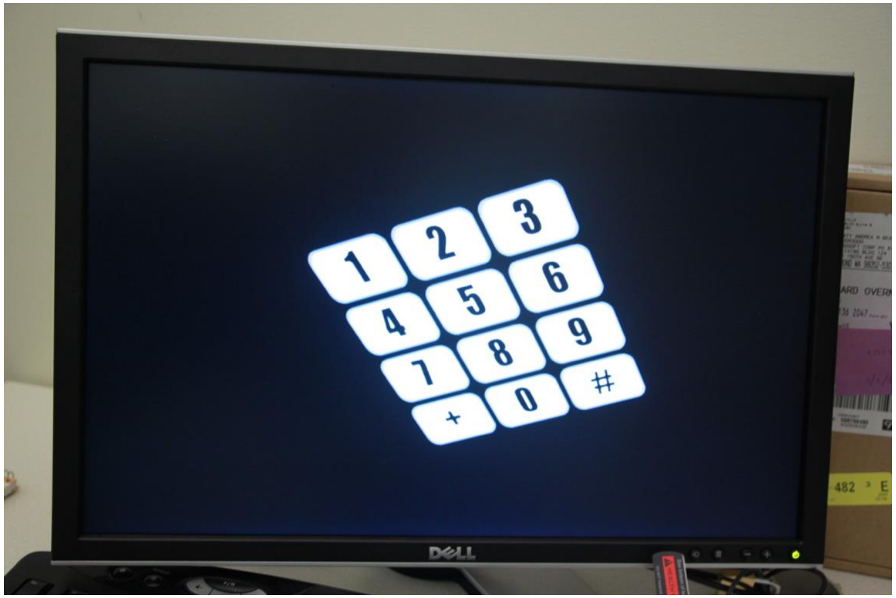
- Take only the ends of objects with physical extent ("fingertips")
- Detect contact ("click")
- Refinement of position while clicked ("drag")

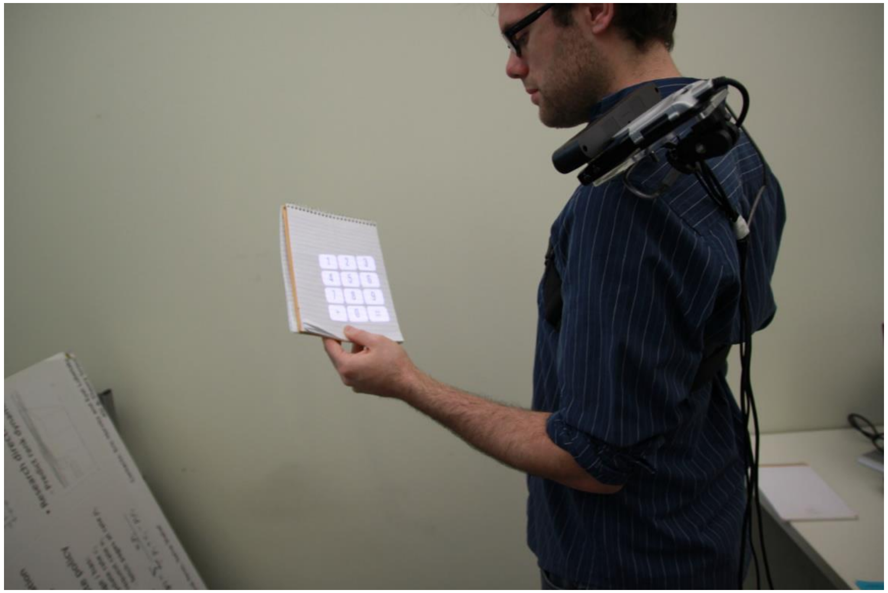




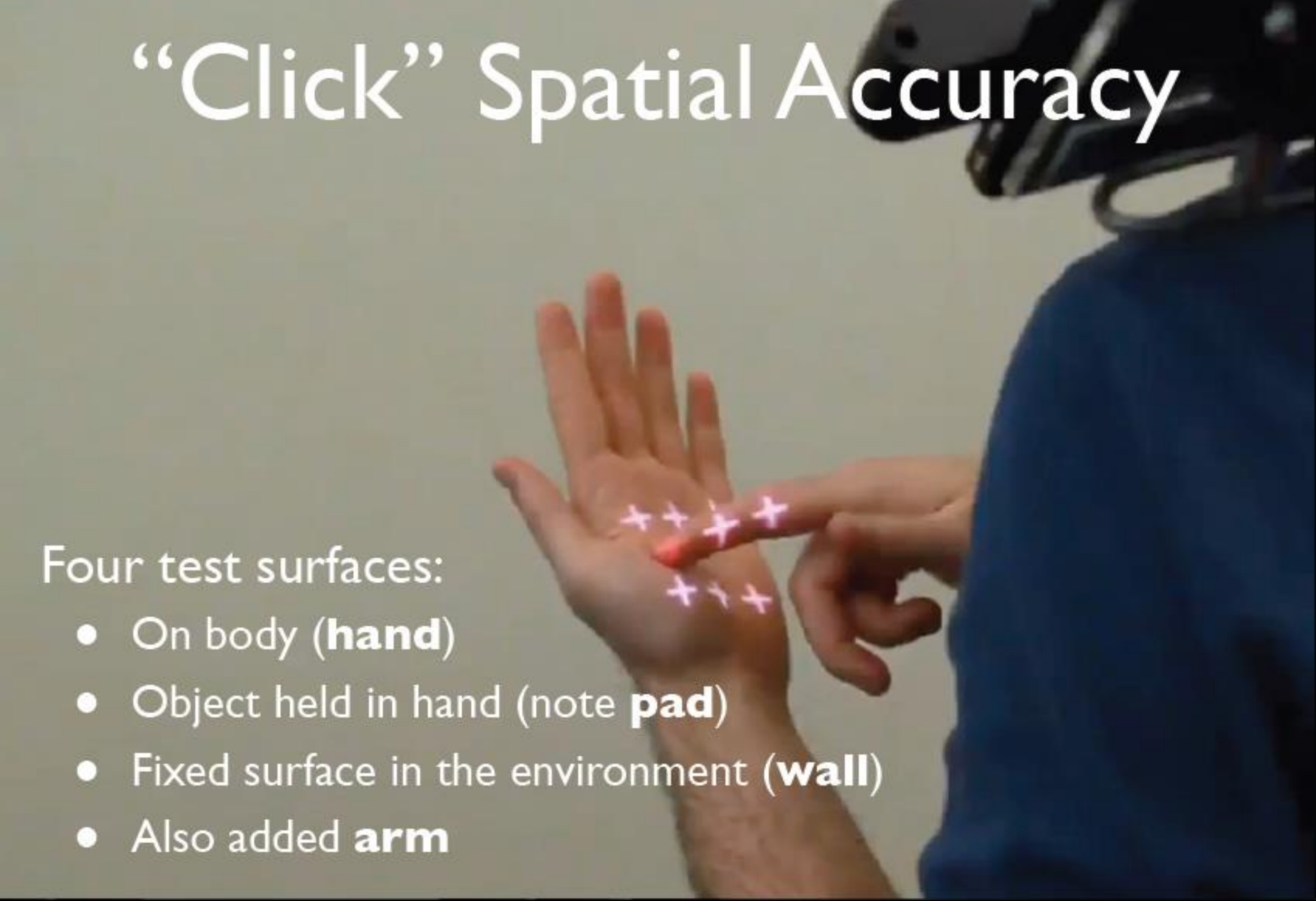
Harrison, Benko, and Wilson, ACM UIST 2011

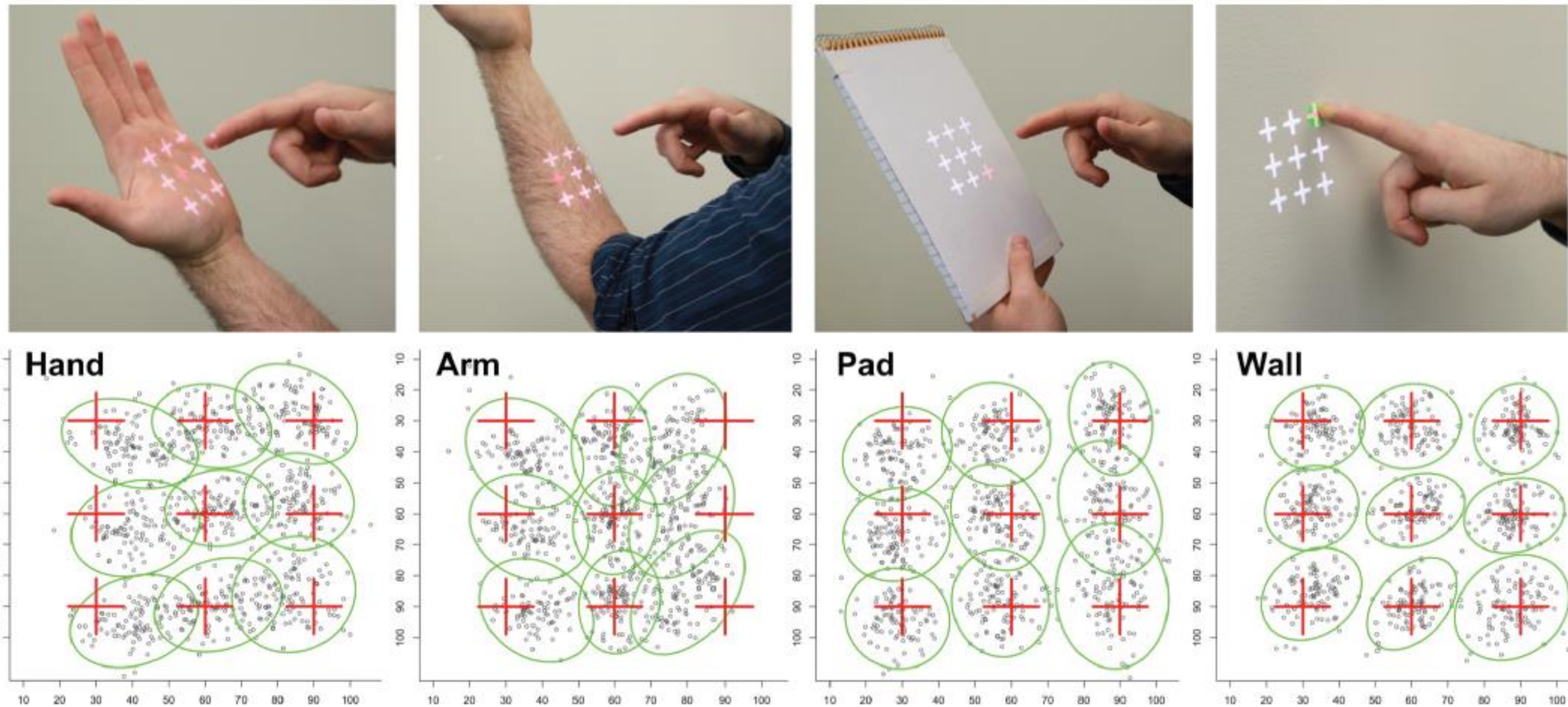



1 2 3
4 5 6
7 8 9
+ 0 #
DELL



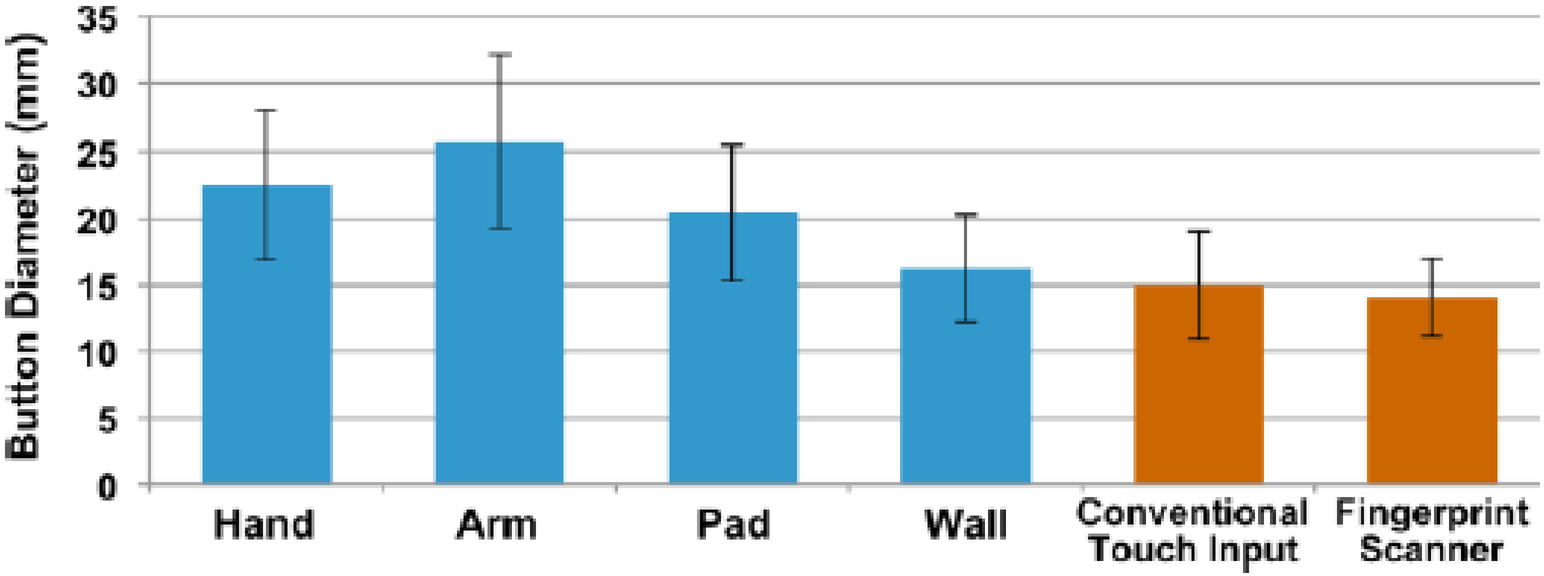

1 2 3
4 5 6
7 8 9
+ 0 #
DELL


"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

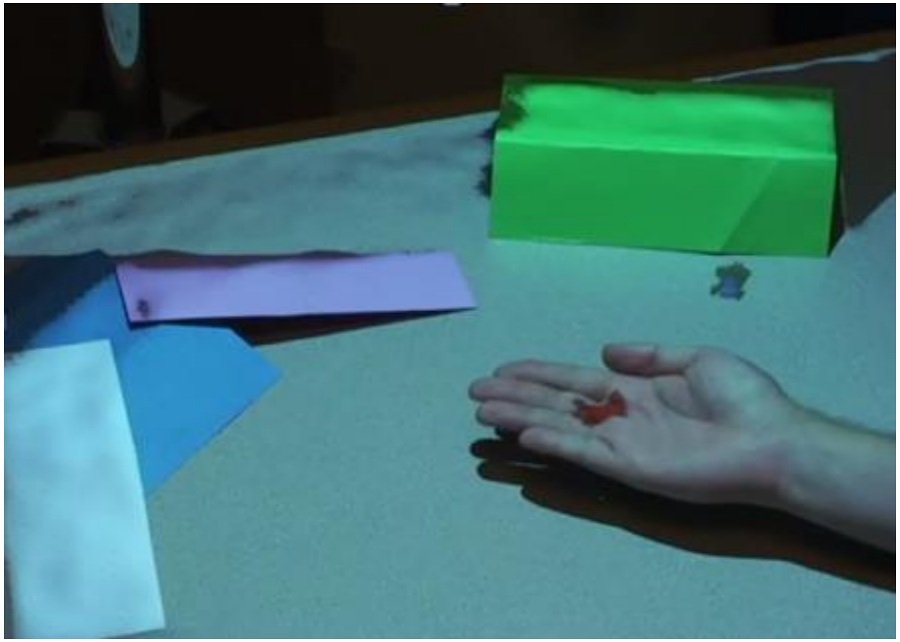


Holz and Baudisch - CHI '10

With 0.5s timeout rejection ~ 98.9% click accuracy

INTERACTING WITH 3D OBJECTS

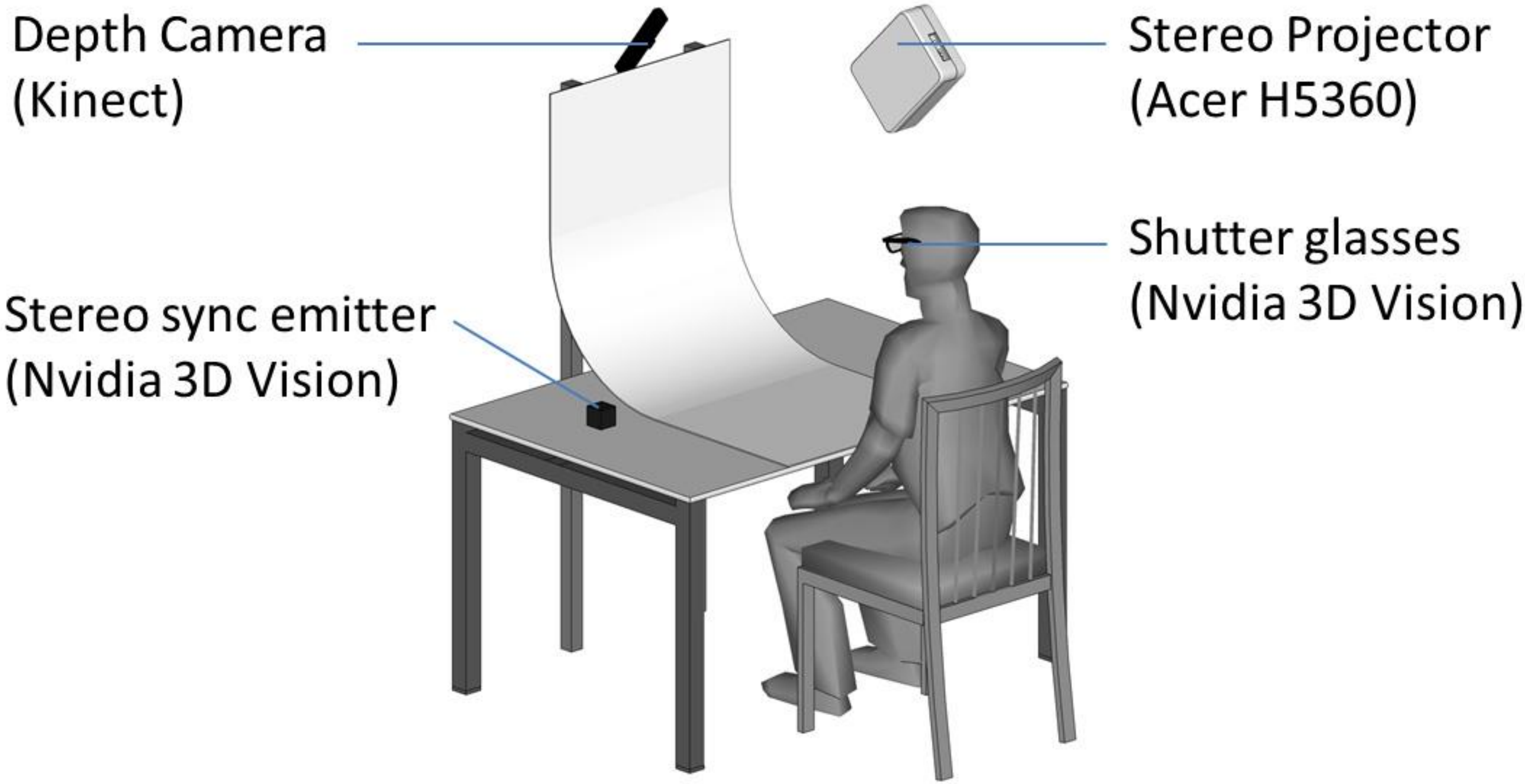
Previous approaches were 2D



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

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

MirageTable



MirageTable

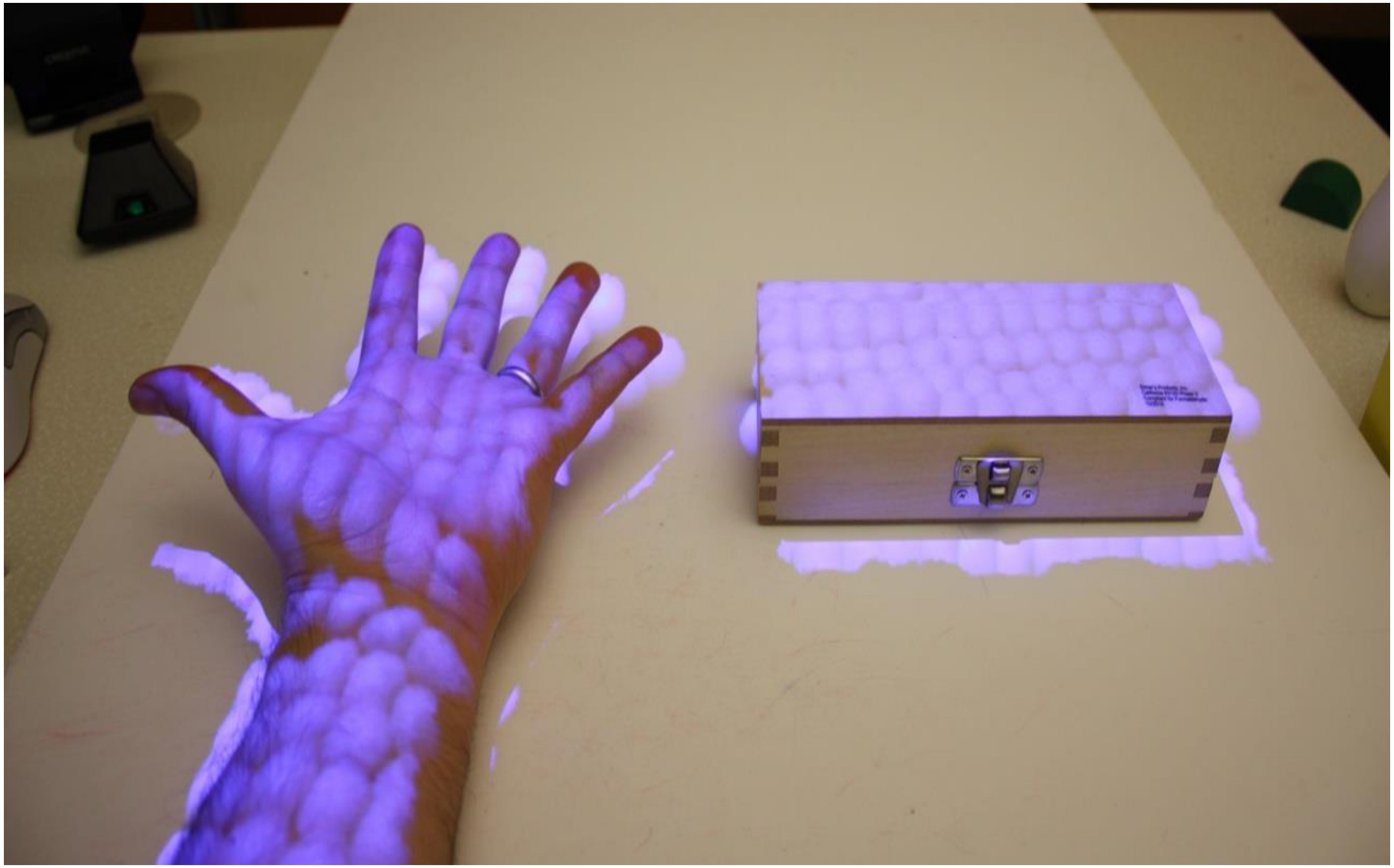
Benko, Jota, Wilson, CHI 2012

Depth camera

↓

If you know the geometry of the world, you can simulate virtual object's physical behavior and 3D appearance.

To hold an object – use particle proxies



Simulating physics is not easy

PhysX® by NVIDIA

newton™ GAME DYNAMICS

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

If you know the geometry of the world, you can simulate virtual object's physical behavior and 3D appearance.

3D Perception

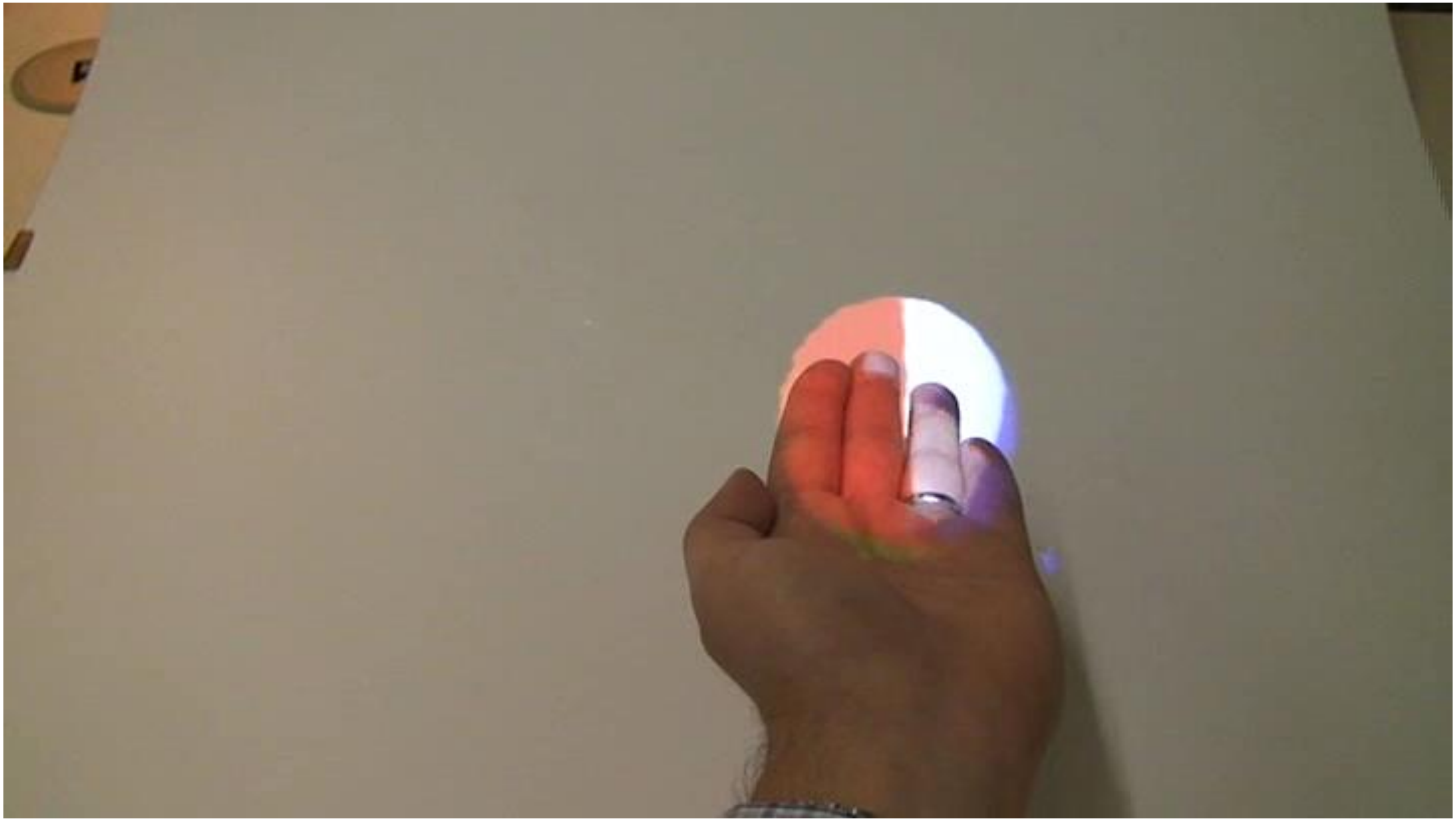
Many cues:

- Size
- Occlusions
- Shadows
- Motion parallax
- Stereo
- Eye focus and vergence

Can correctly simulate **if** you know:

- The geometry of the scene
- User's view point and gaze

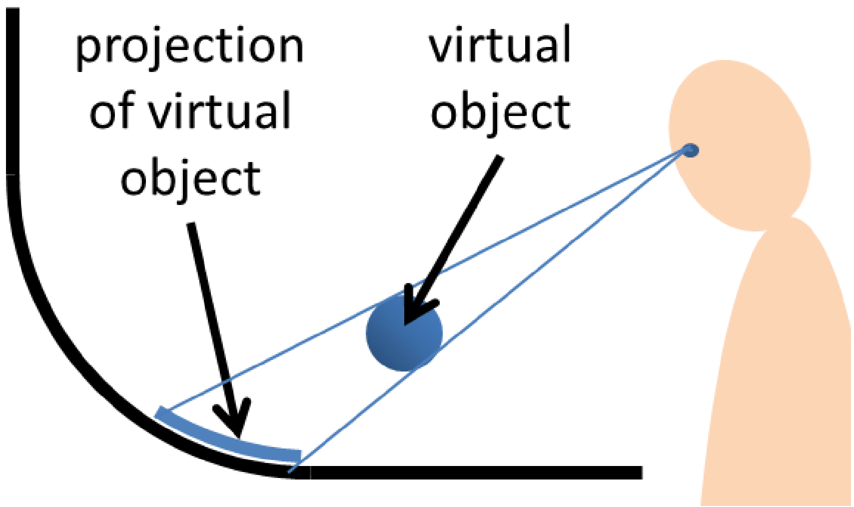
3D in your hand



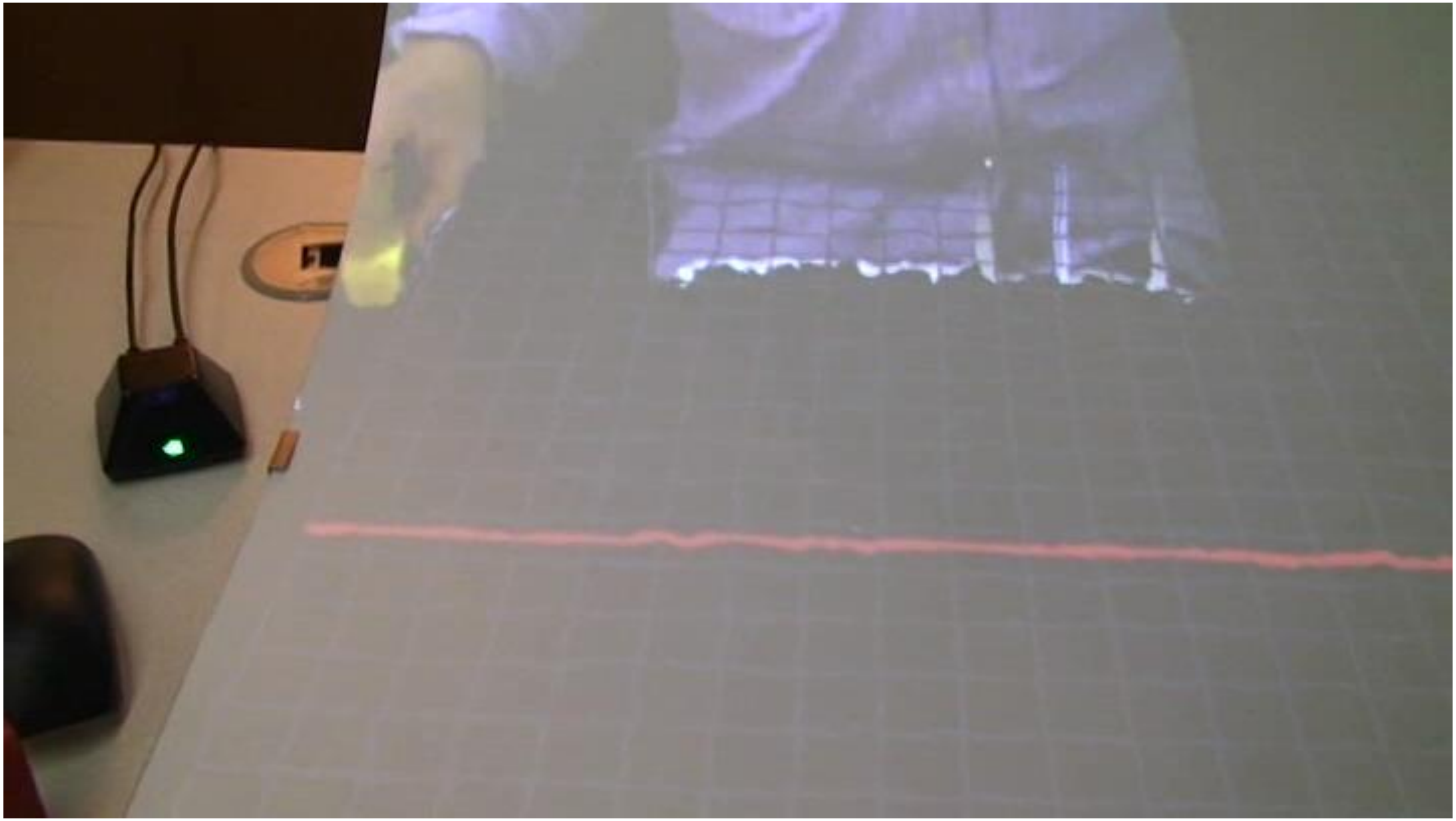
Benko, Jota, Wilson, 2012

Projective texturing

Background only



Example Application Scenarios



Benko, Jota, Wilson, CHI 2012

Summary

1. Interactivity everywhere
2. Room and body as display surfaces
3. Touch and 3D interactions
4. Preserve the analog feel of interactions

My collaborators



Andy Wilson

Microsoft Research

Chris Harrison

Carnegie Mellon University

Ricardo Costa Jota

Instituto Superior Tecnico, Lisabon

Hrvoje Benko

benko@microsoft.com

http://research.microsoft.com/~benko

Microsoft®

Microsoft
Research