Microsoft® Research

FacultySummit 2011

Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS

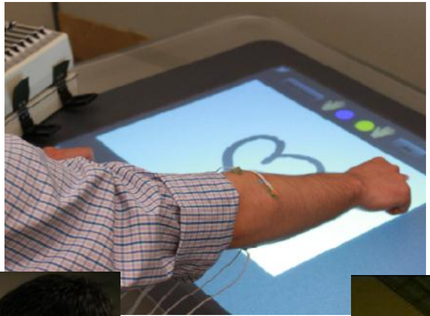
Microsoft® Research

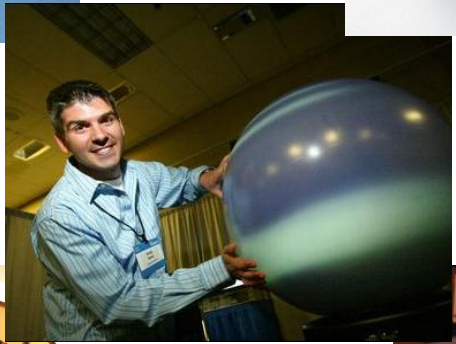
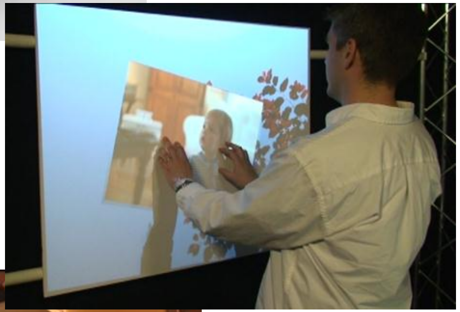
FacultySummit 2011

Cartagena, Colombia | May 18-20 | In partnership with COLCIENCIAS

High-Fidelity Augmented Reality Interactions

Hrvoje Benko
Researcher, MSR Redmond



New generation of interfaces

Instead of interacting through indirect input devices (mice and keyboard), the user is interacting directly with the content.

Direct un-instrumented interaction

Content is the interface

Surface computing

Kinect

New generation of interfaces

Direct un-instrumented interaction.

Content is the interface.

New generation of interfaces

Bridge the gap between
"real" and "virtual" worlds...


… but still confined to
the rectangular screen!

An opportunity…

Depth camera + Projector

Enable interactivity on any available surface and between surfaces.

MicroMotoCross

Wilson, 2007

Augmented reality

Spatial

"Deviceless"

High-fidelity

Depth Sensing Cameras

Depth sensing cameras

Color + depth per pixel: RGBZ

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



Three basic types

- Stereo
- Time of flight
- Structured light

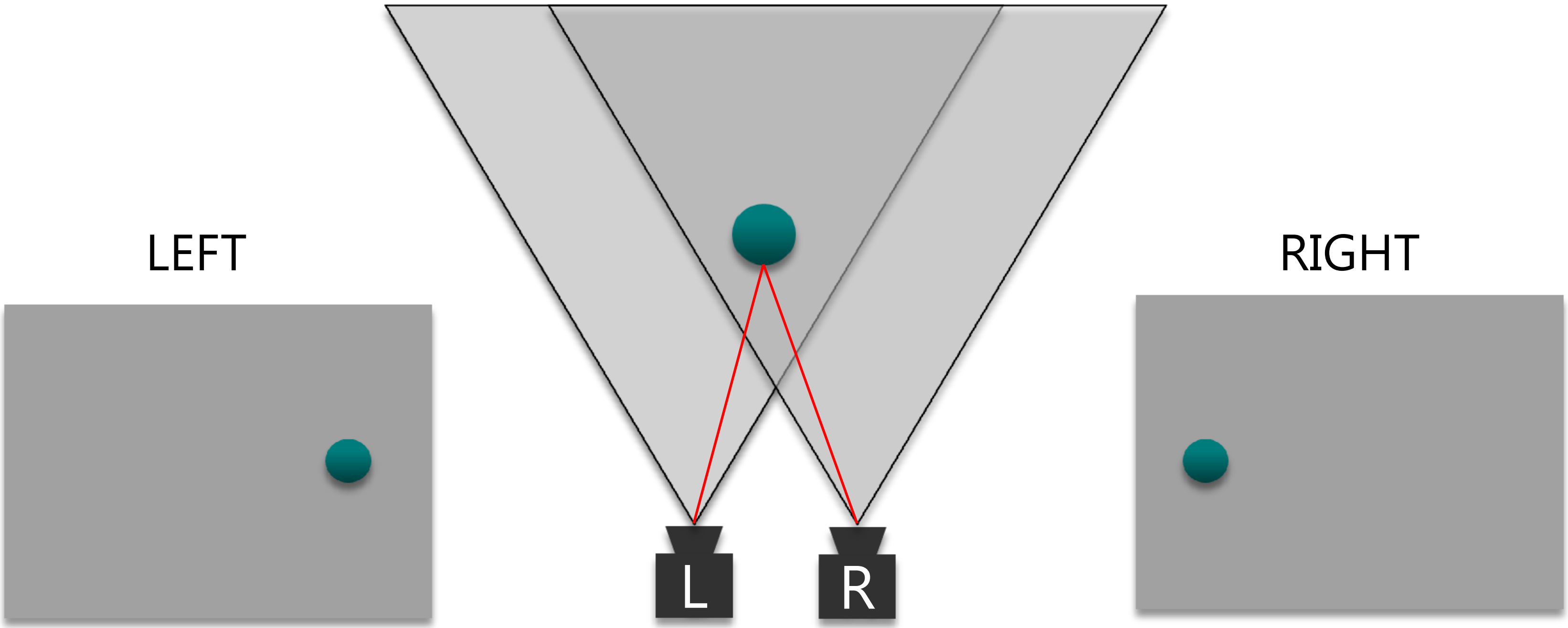
Correlation-based stereo cameras

Binocular disparity

TZYX http://www.tyzx.com/

Point Grey Research http://www.ptgrey.com

Correlation-based stereo



Stereo drawbacks

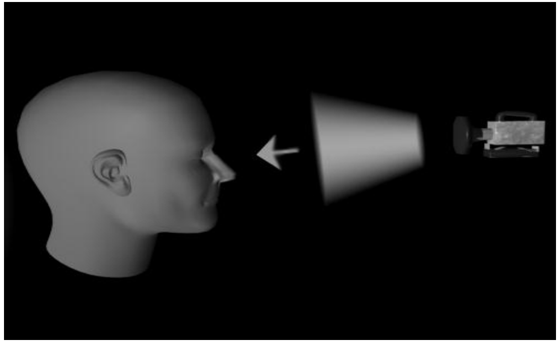
- Requires good texture to perform matching
- Computationally intensive
- Fine calibration required
- Occlusion boundaries
- Naïve algorithm very noisy

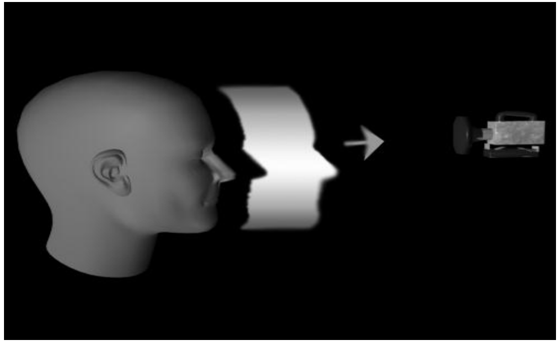
Time of flight cameras

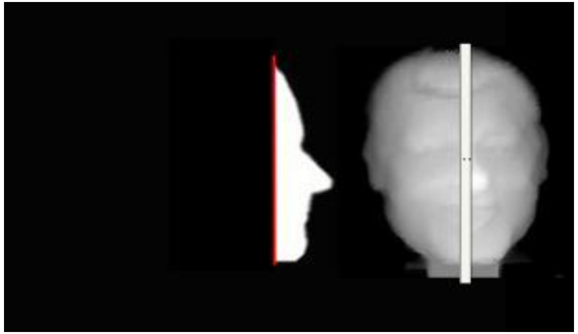
3DV ZSense

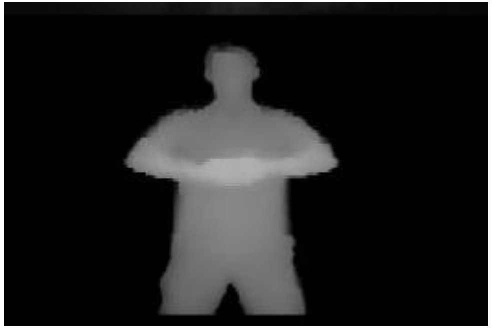
3DV, Canesta (no-longer public)
PMD Technologies http://www.PMDTec.com
Mesa Technologies http://www.mesa-imaging.ch

Time of flight measurement









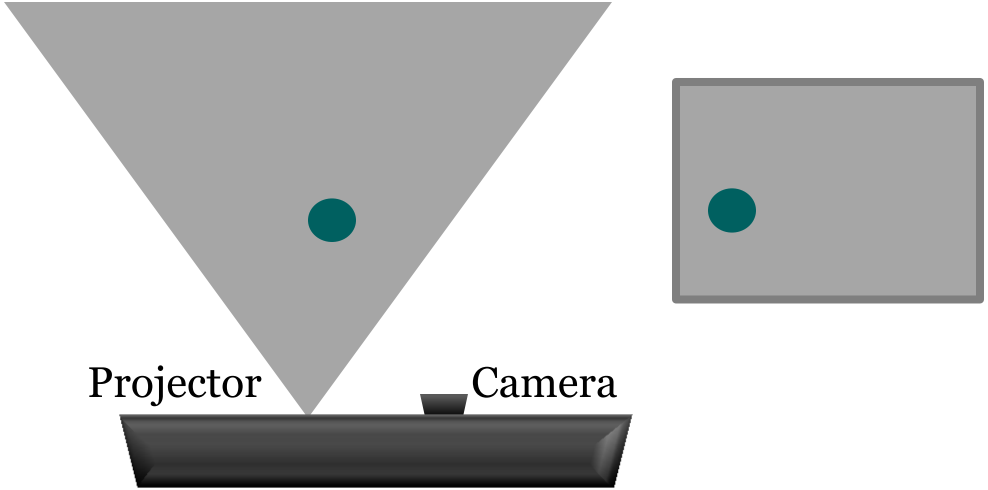
Structured light depth cameras

KINECT™ for XBOX 360.

PrimeSense http://www.primesense.com
Microsoft Kinect http://www.microsoft.com/kinect

Structured light (infrared)

Depth by binocular disparity



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

Kinect depth camera

- Per-pixel depth (mm)
- PrimeSense reference design
- Field of View 58° H, 45° V, 70° D
- Depth image size VGA (640x480)
- Spatial x/y resolution (@ 2m distance from sensor) 3mm
- Depth z resolution (@ 2m distance from sensor) 1cm
- Operation range 0.8m - 3.5m

- Best part – It is affordable - $150

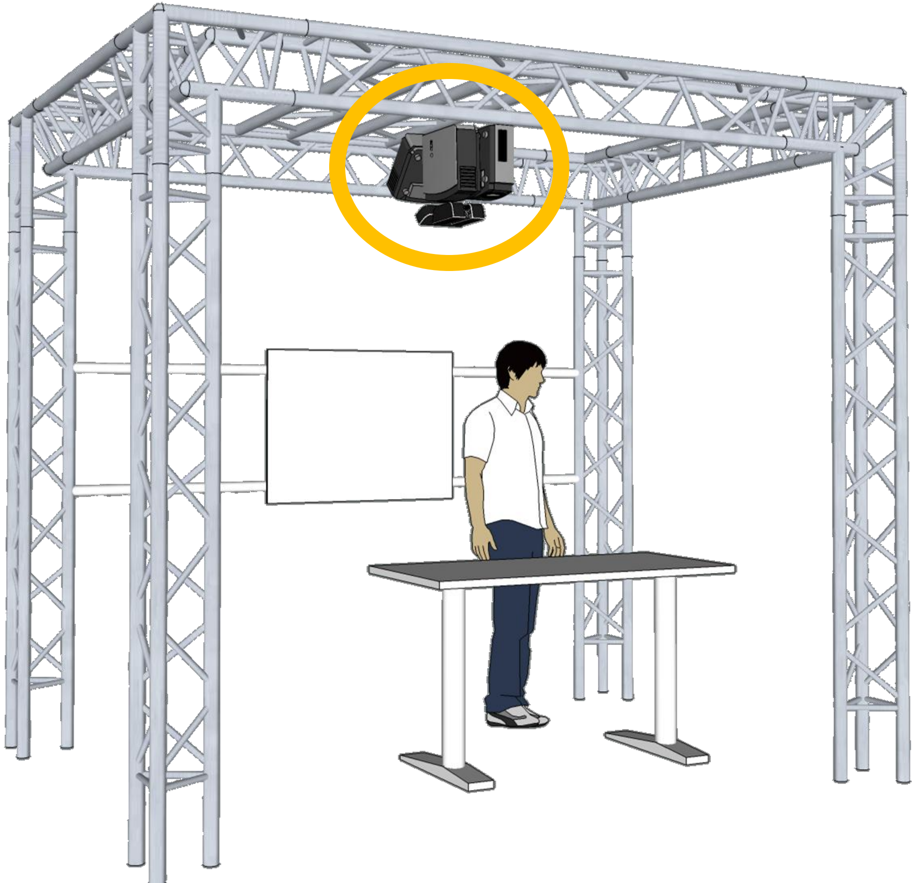
Why sense with depth cameras?

Requires no instrumentation of the surface/environment.

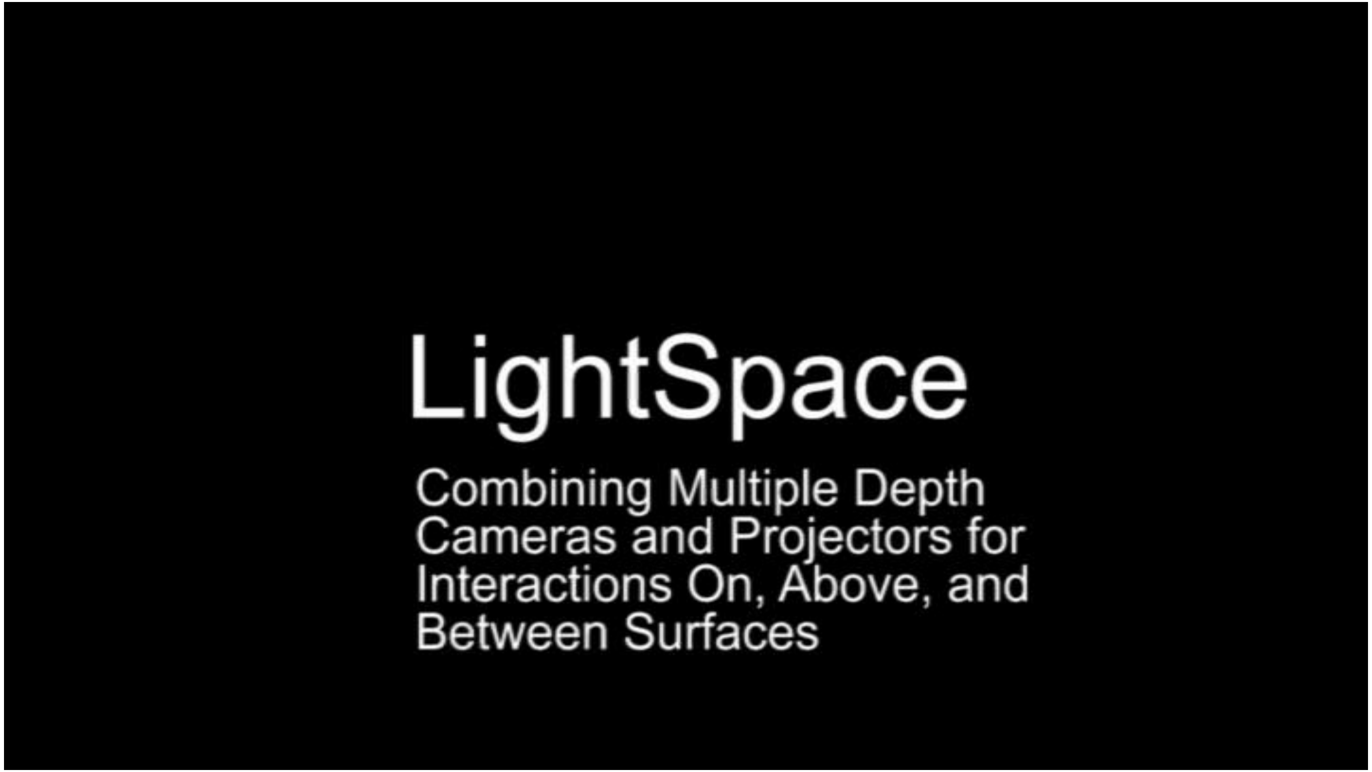
Easier understanding of physical objects in space.

Enabling interactivity everywhere

LightSpace



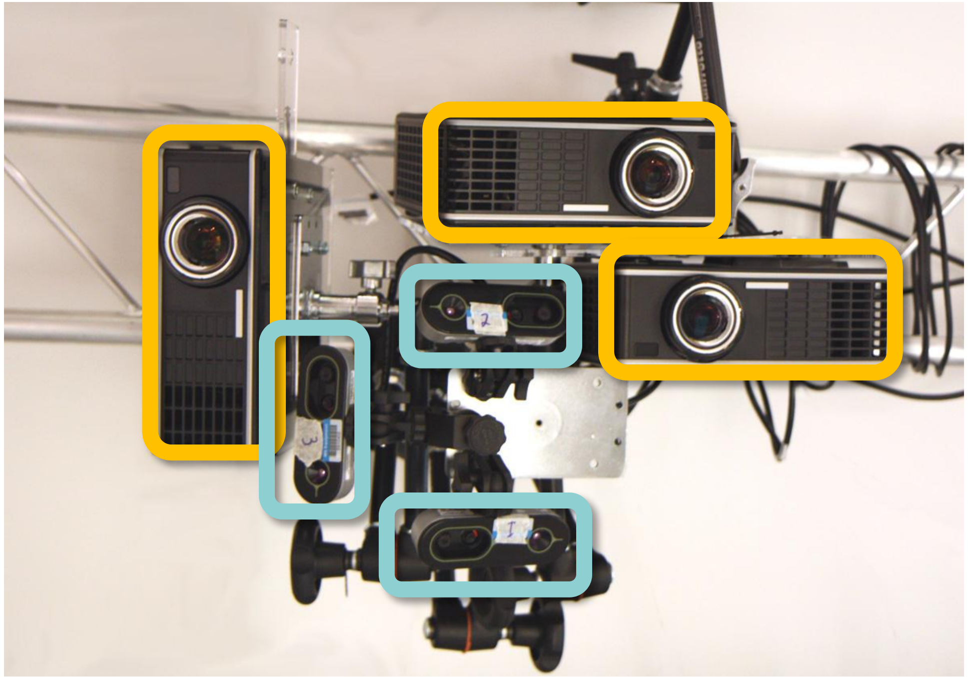
LightSpace



Wilson & Benko, 2010

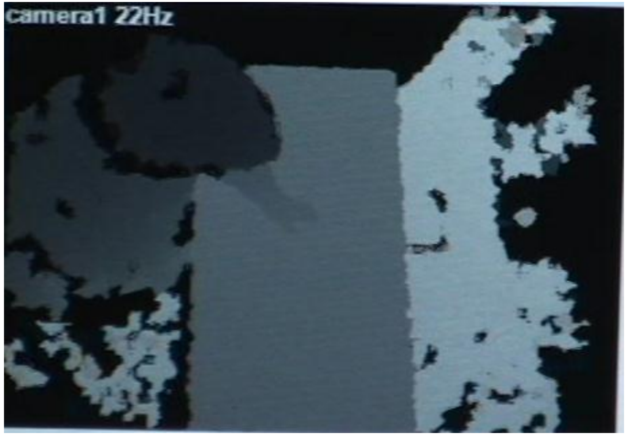
LightSpace Implementation

Projectors



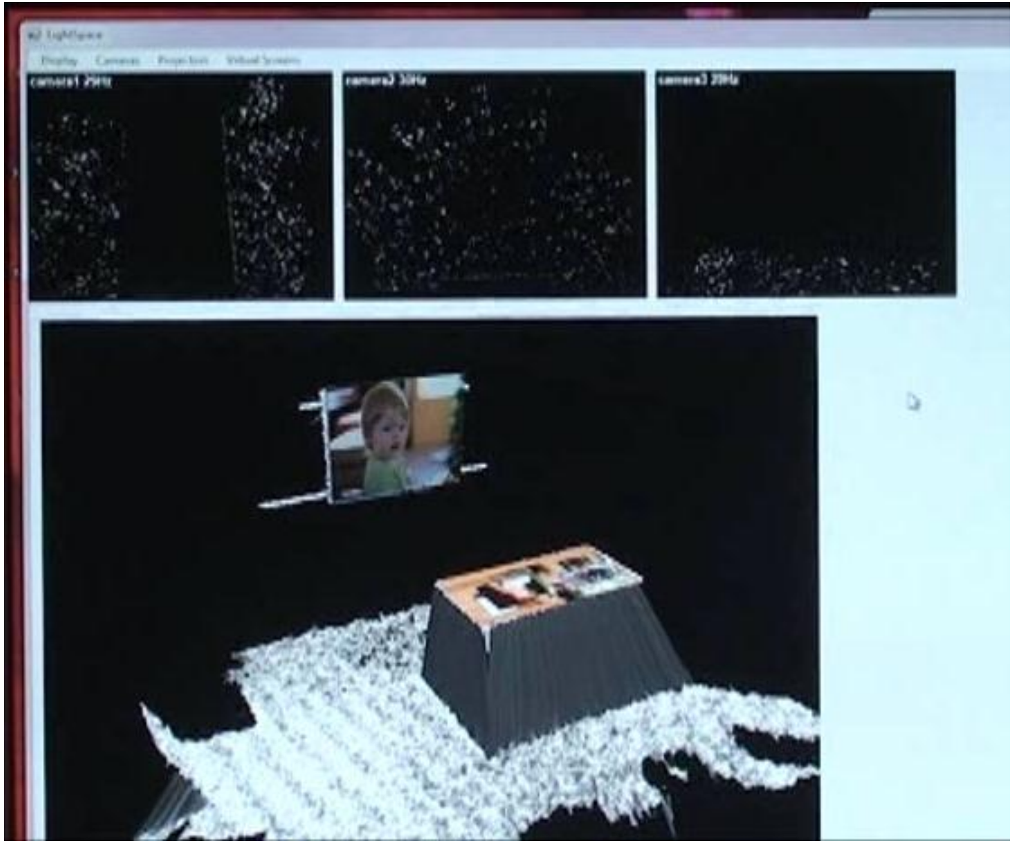
PrimeSense Depth Cameras

PrimeSense depth cameras

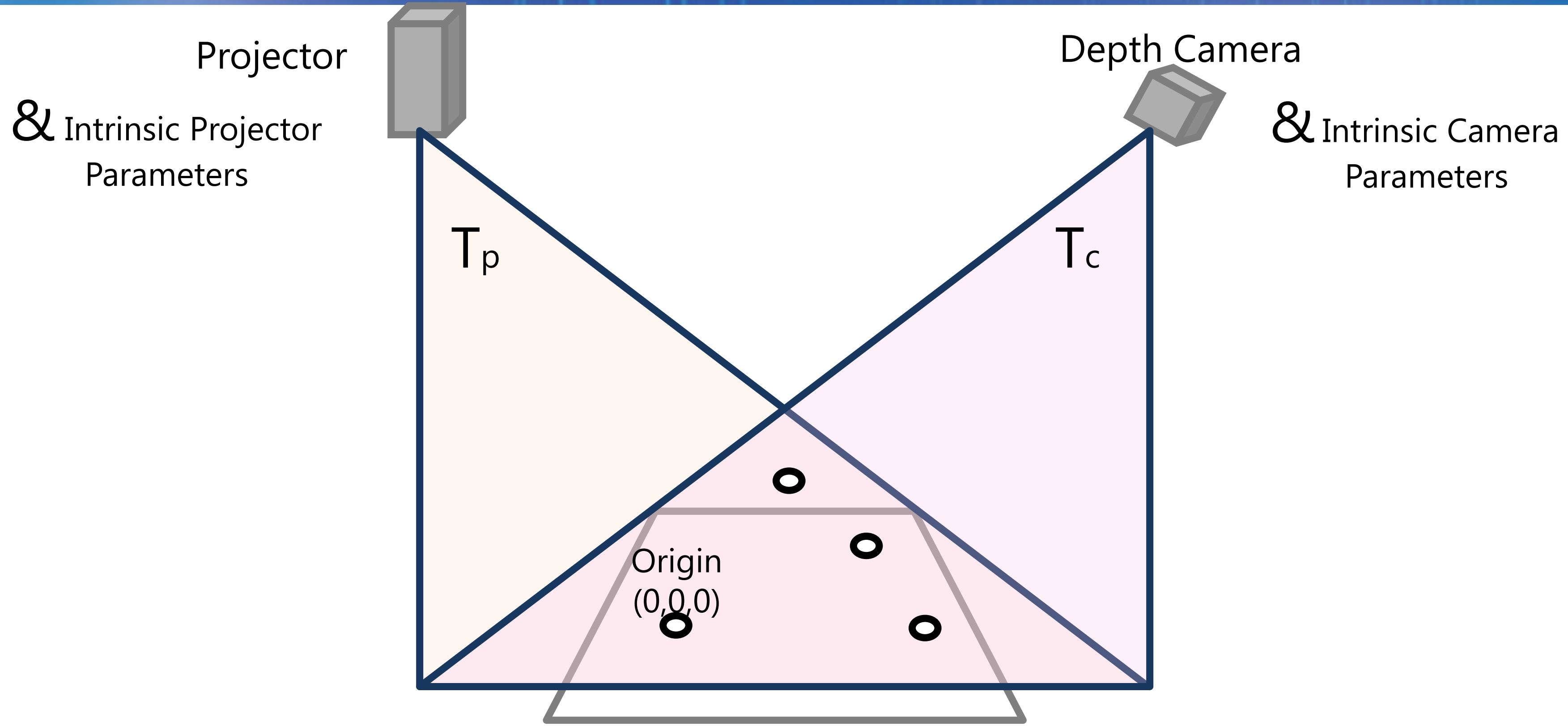


320x240 @ 30Hz
Depth from projected structured light
Small overlapping areas
Extended space coverage

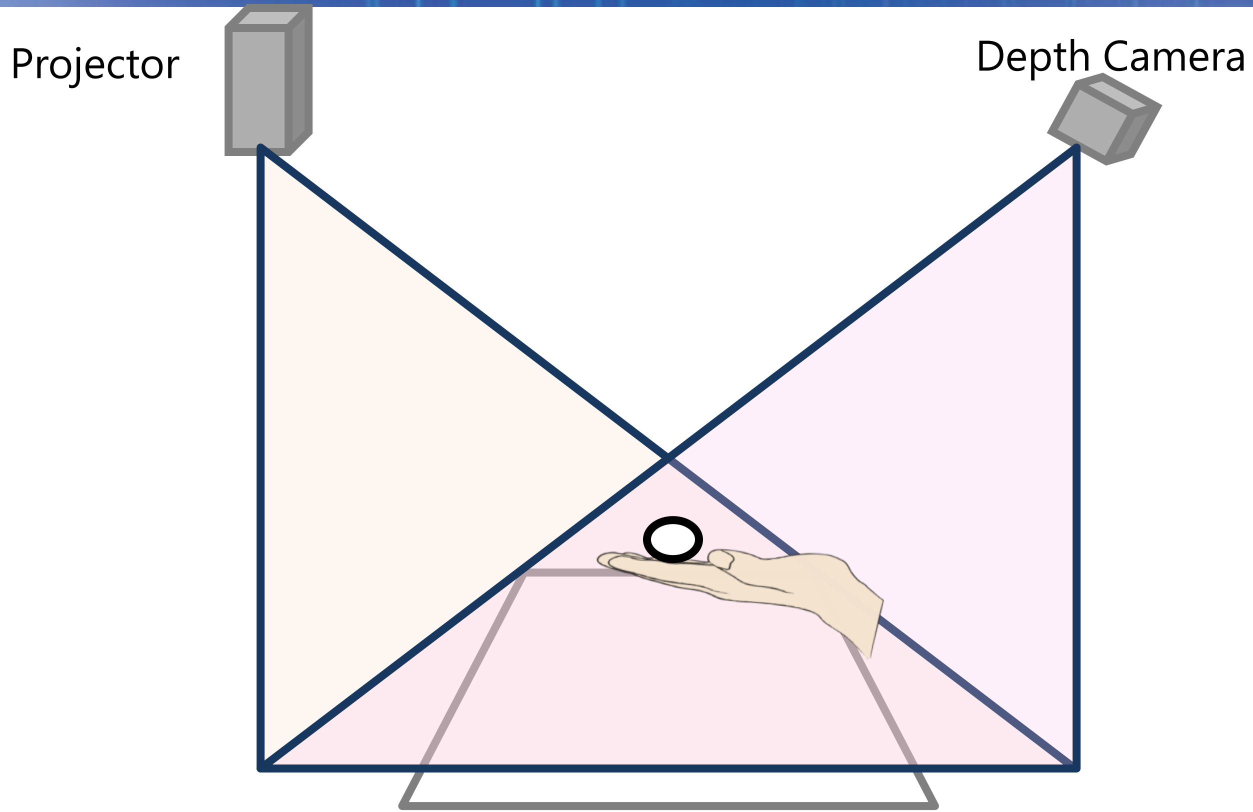
Unified 3D Space



Camera & projector calibration



Camera & projector calibration



LightSpace authoring

All in real world coordinates.

Irrespective of “which” depth camera.

Irrespective of “which” projector.

Supporting rich analog interactions

Skeleton tracking (Kinect)

Our approach

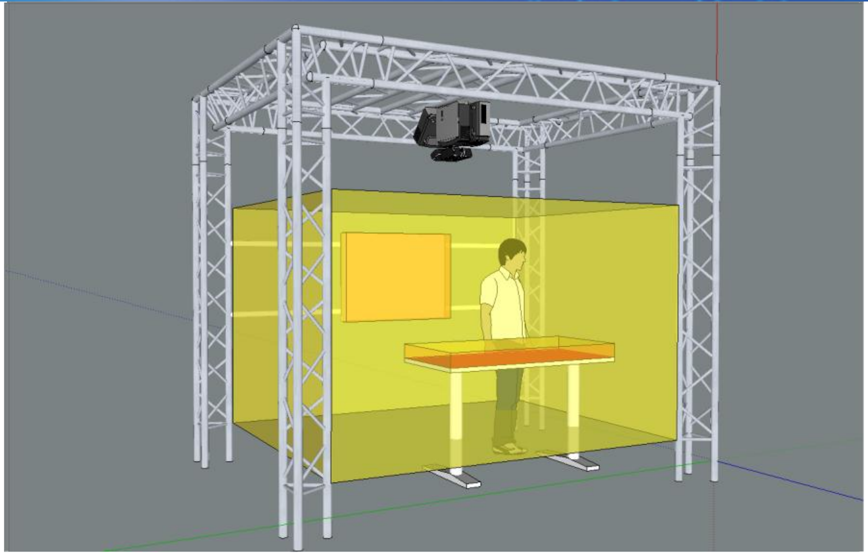
Use the full 3D mesh.

Preserve the analog feel through physics-like behaviors.

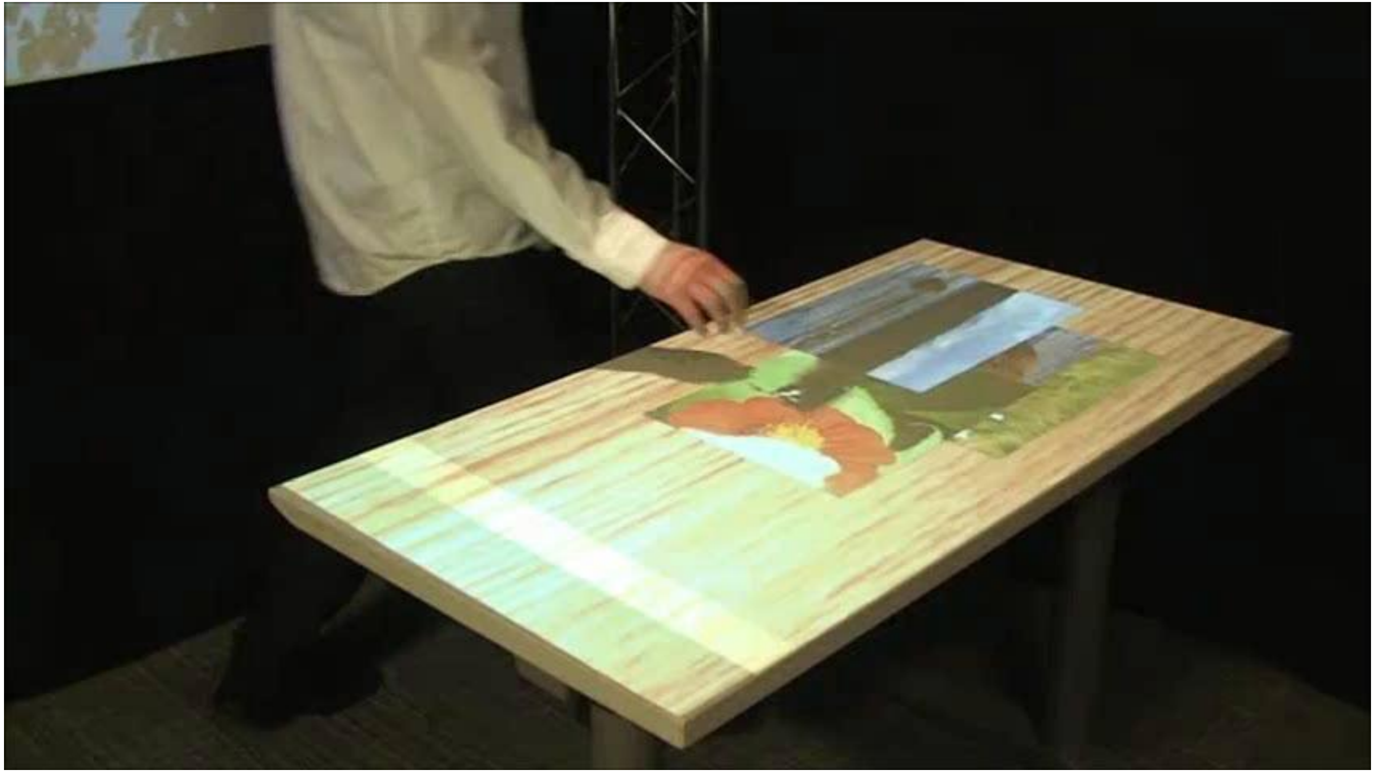
Reduce the 3D reasoning to 2D projections.

Pseudo-physics behavior

Virtual depth cameras



Simulating virtual surfaces

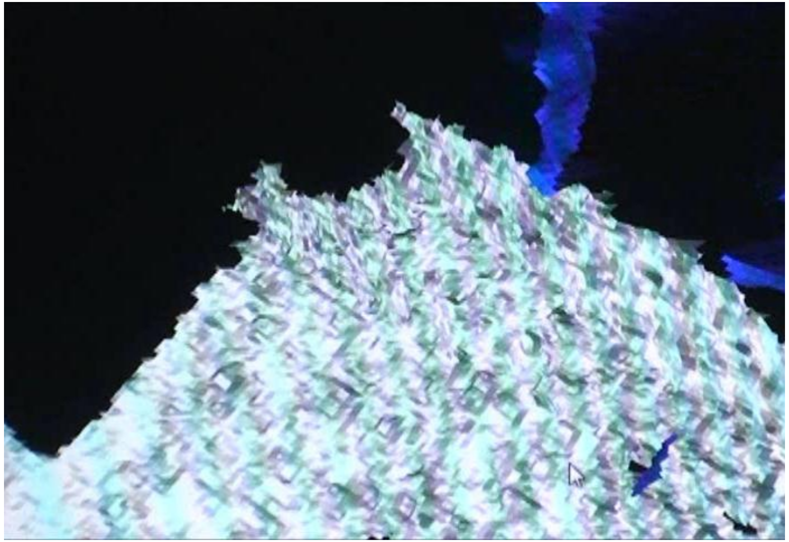


Through-body connections

Physical connectivity

Spatial widgets

User-aware, on-demand spatial menu



What is missing?

LightSpace

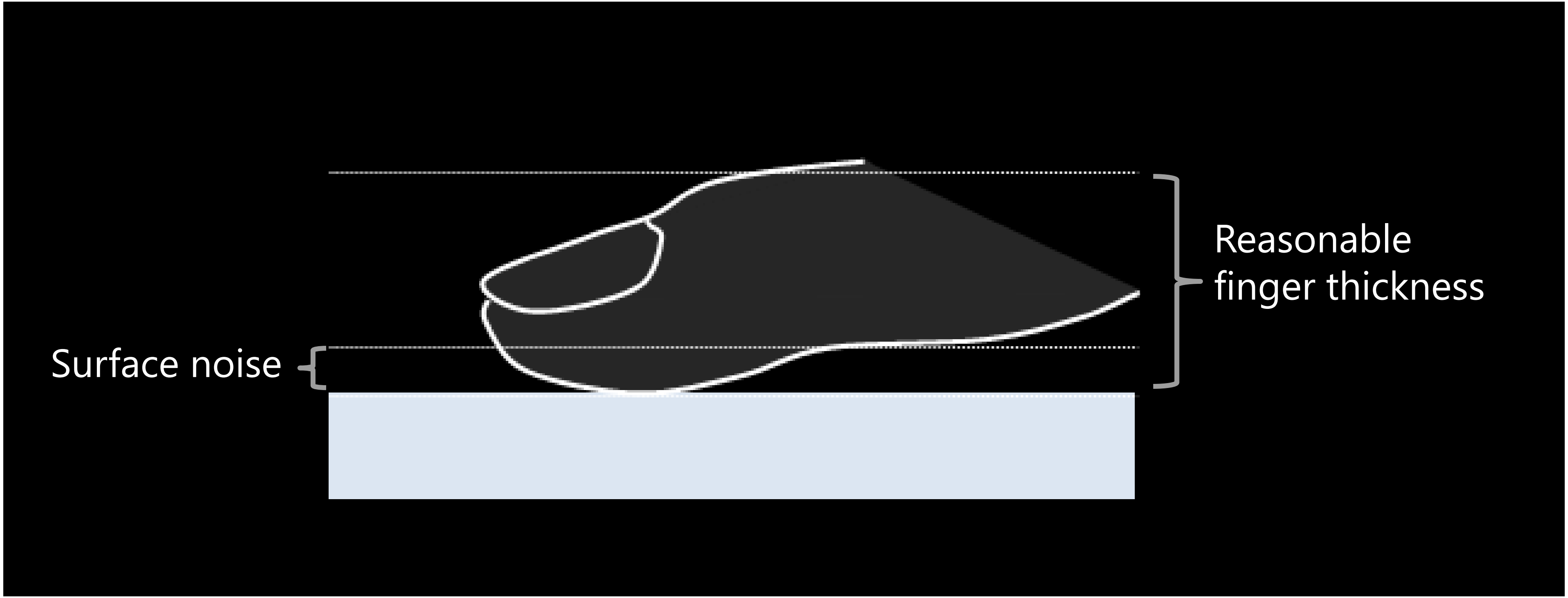
- "Touches" are hand blobs
- All objects are 2D
- Very coarse manipulations

Ideally

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

Touch on every surface

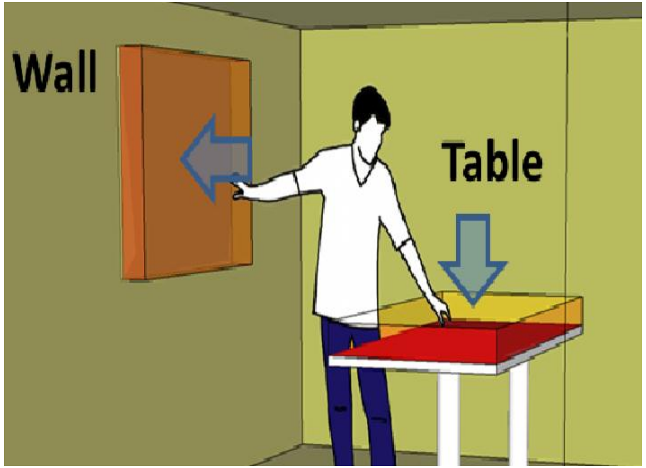
Problem of two thresholds



How to get surface distance?

Analytically

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



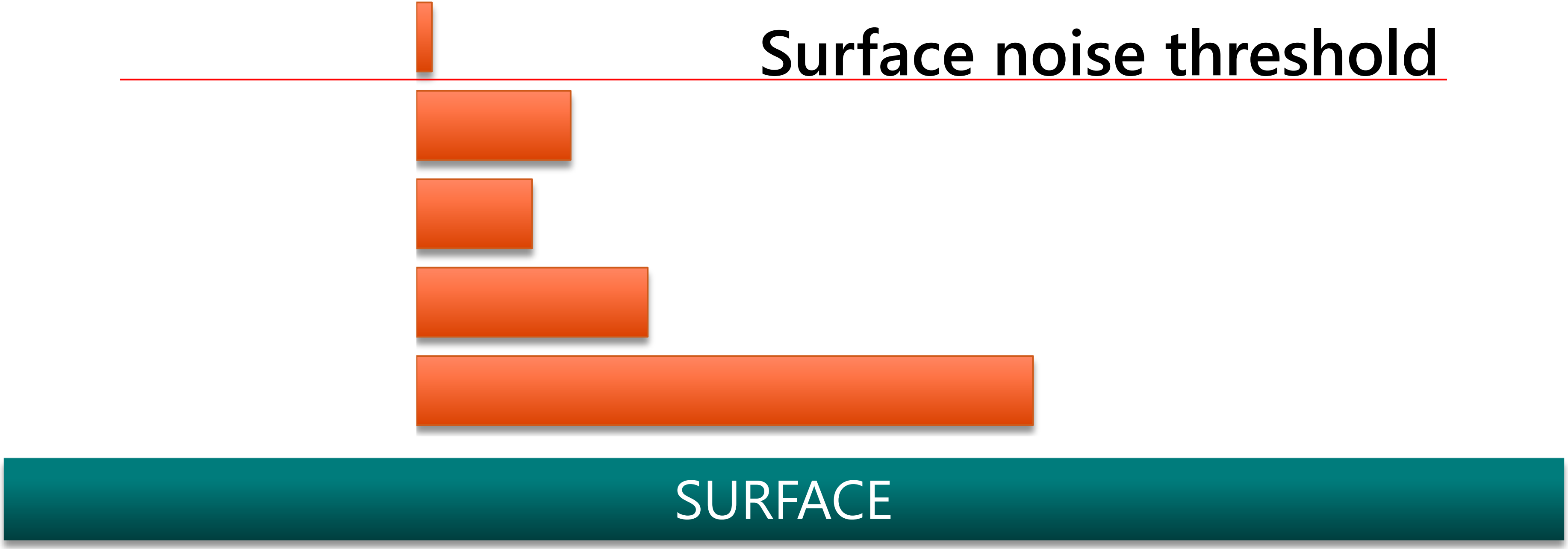
How to get surface distance?

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

Modeling the surface

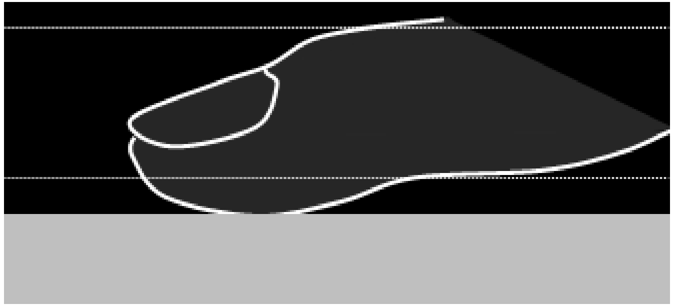
Build a surface histogram at every pixel.



Setting reasonable finger thickness

Must make some assumption about anthropometry, posture, and noise.

How good can you get?



Camera above surface	0.75m	1.5m
Finger threshold	14mm	30mm
Surface noise	3mm	6mm

KinectTouch

Wilson 2010

But these are all static surfaces

How to allow touch on any (dynamic) surface?

- Dynamic surface calibration
- Tracking high-level constructs such as finger posture, 3D shape
  - Take only the ends of objects with physical extent (“fingertips”)
  - Refinement of position

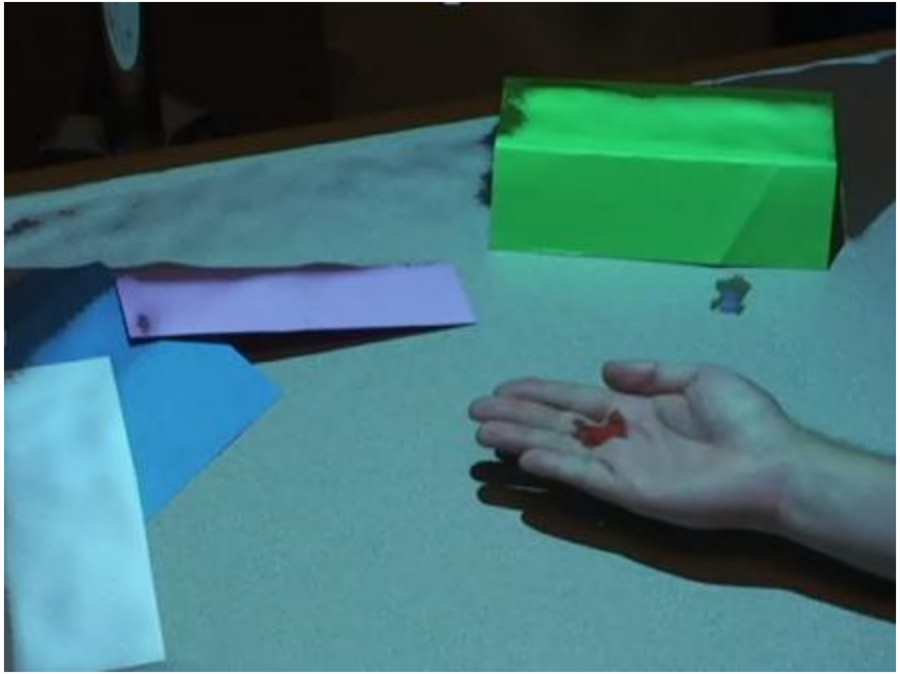
Depth camera touch sensing is almost as good as conventional touch screen technology!

Works on any surface!

(curved, flexible, deformable, flat…)

Interacting with 3D objects

Previous approaches were 2D



Micromotocross

LightSpace

Can one hold a virtual 3D object in their hand?

And manipulate it using the full dexterity of your hand?

If you know the geometry of the world, you should be able to simulate physical behaviors.

Problems with physics and depth cameras

Dynamic meshes are difficult

- Rarely supported in physics packages

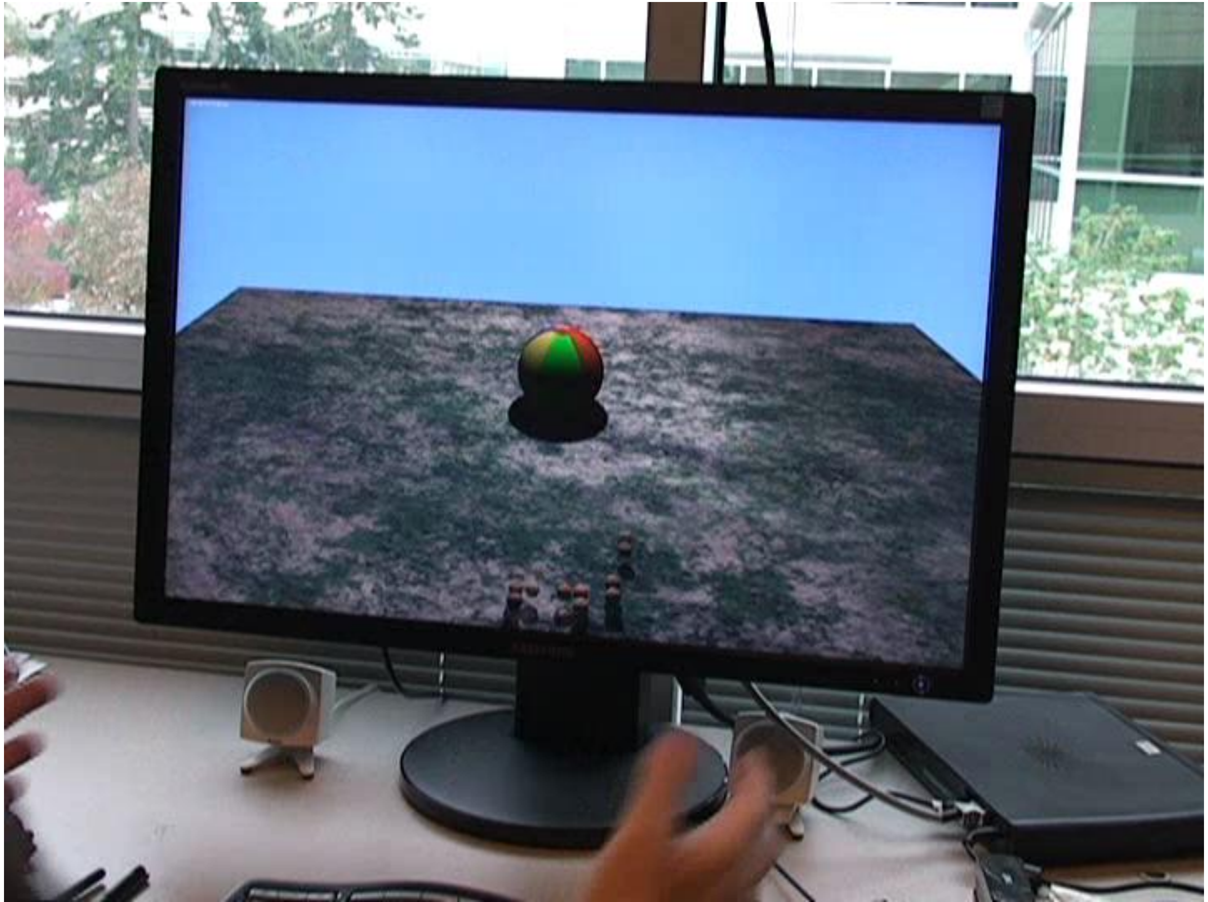
No lateral forces!

- Can’t place torque on an object

Penetration is handled badly

- Can’t grasp an object with two fingers

Particle proxy representations



Wilson 2007

But can you see 3D in your hand?

3D perception

Many cues:

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

Can correctly simulate **if** you know:

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

Depth camera is ideal for this!

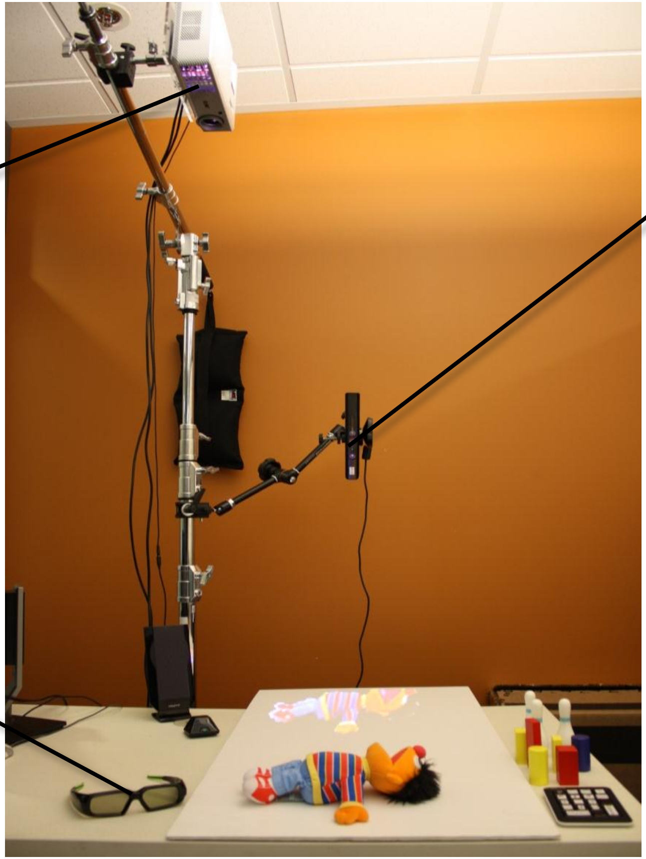
Can easily capture scene geometry

Can easily track user's head

MirageBlocks

3D Projector
(Acer H5360)

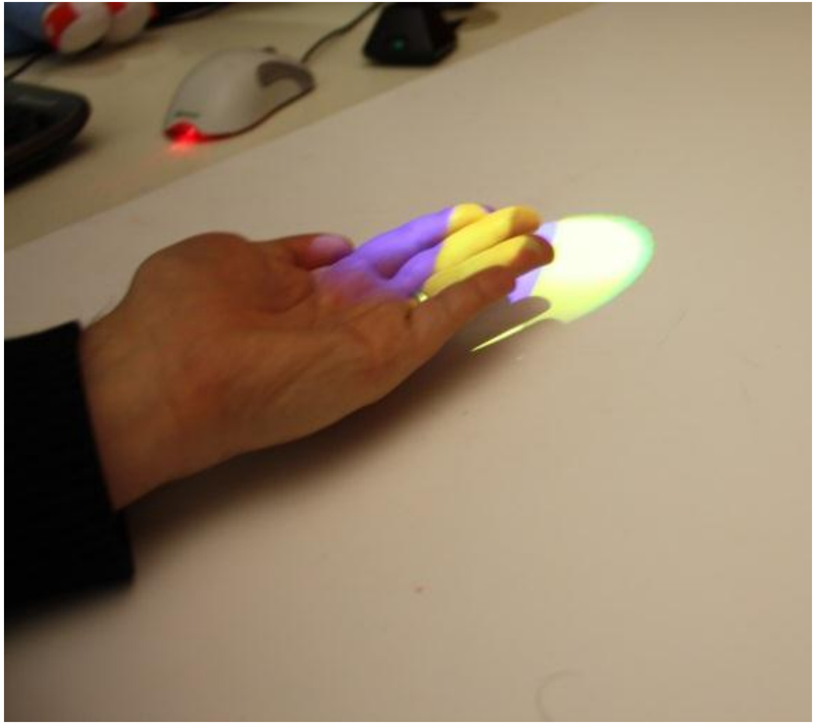
Shutter Glasses
(Nvidia 3D Vision)



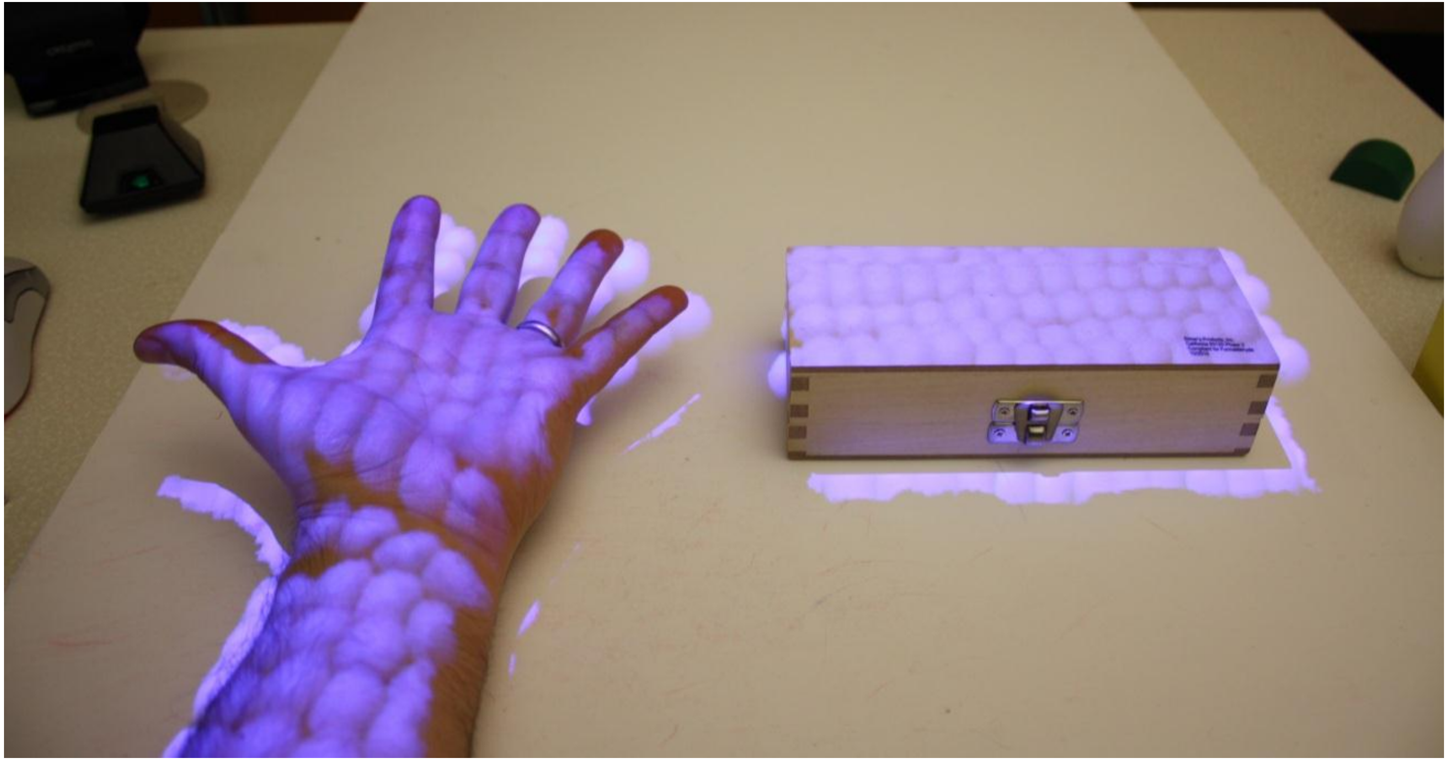
Depth Camera
(Kinect)

Benko, Costa, and Wilson, 2011

A single user experience!



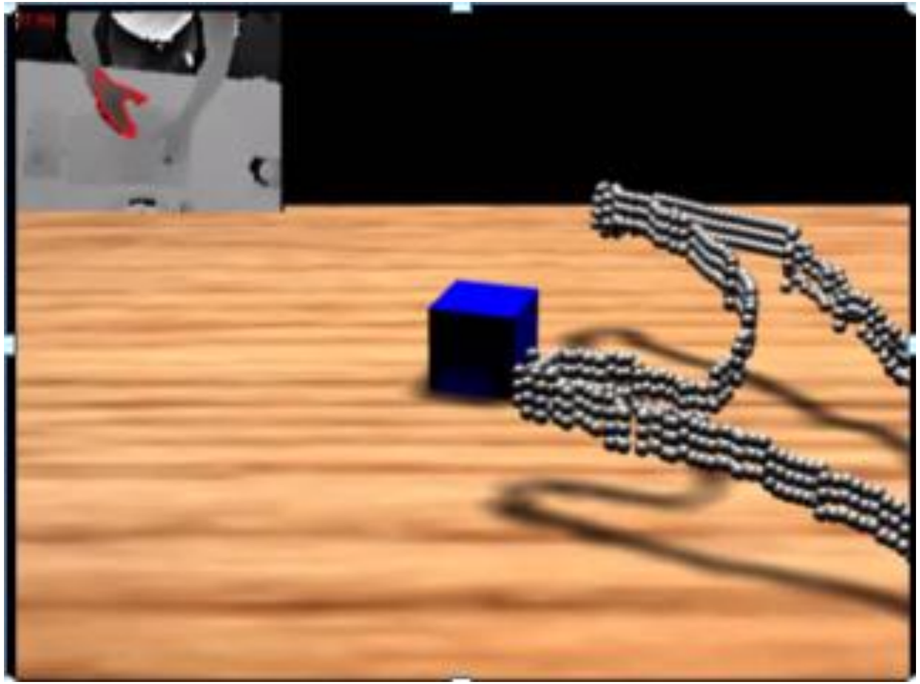
Particle proxies



MirageBlocks

Next: Grabbing

Very hard problem – Working on it!



Summary

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

Come to try it yourself!

MirageBlocks demo

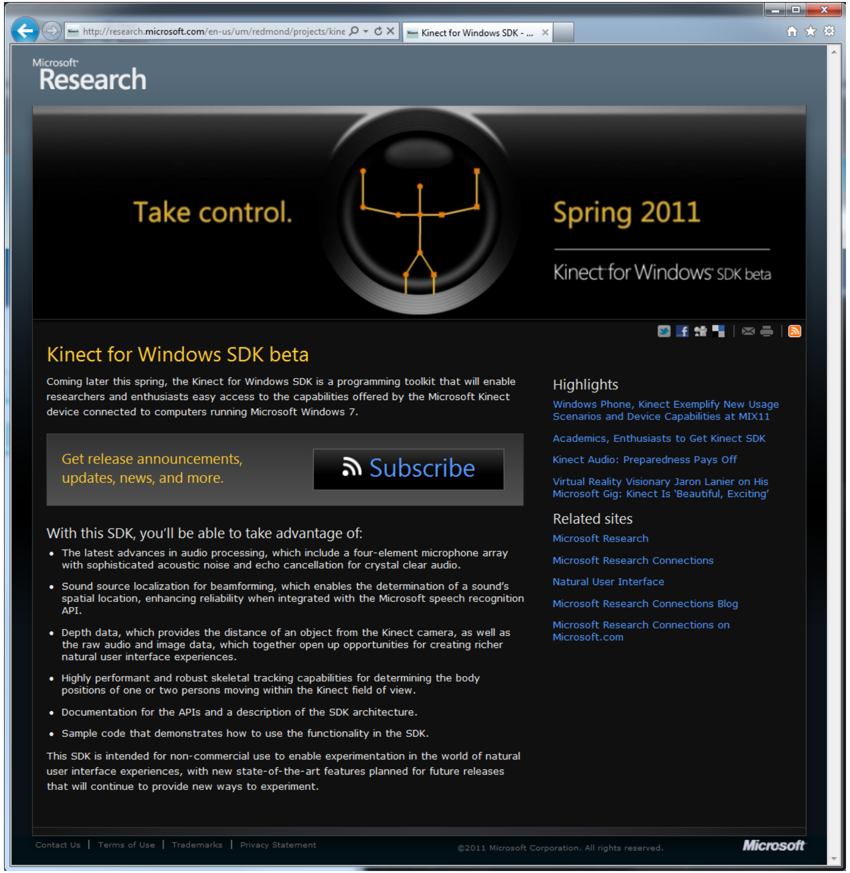
Friday 10am – 1pm

Resources to consider

Resources

Kinect for Windows SDK

- [http://research.microsoft.com/en-us/um/redmond/projects/**kinectsdk**](http://research.microsoft.com/en-us/um/redmond/projects/kinectsdk)



Resources

NVIDIA PhysX SDK

- http://developer.nvidia.com/physx-downloads
- http://physxdotnet.codeplex.com/ (.NET wrappers)

Newton Physics Game Engine

- http://newtondynamics.com/forum/newton.php

Resources

NVIDIA 3D Vision

- http://www.nvidia.com/object/3d-vision-main.html

DLP Link

- http://www.dlp.com/projector/dlp-innovations/dlp-link.aspx
- http://www.xpand.me/ (3D glasses)

My collaborators

Andy Wilson

Chris Harrison

Ricardo Costa Jota

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

Microsoft®